Combinatorial theorems in sparse random sets

D. Conlon[*]   W. T. Gowers [†]

Abstract

We develop a new technique that allows us to show in a unified way that many well-known combinatorial theorems, including Turán's theorem, Szemerédi's theorem and Ramsey's theorem, hold almost surely inside sparse random sets. For instance, we extend Turán's theorem to the random setting by showing that for every $\epsilon > 0$ and every positive integer $t \geq 3$ there exists a constant C such that, if G is a random graph on n vertices where each edge is chosen independently with probability at least $Cn^{-2/(t+1)}$, then, with probability tending to 1 as n tends to infinity, every subgraph of G with at least $\left(1 - \frac{1}{t-1} + \epsilon\right) e(G)$ edges contains a copy of K_t. This is sharp up to the constant C. We also show how to prove sparse analogues of structural results, giving two main applications, a stability version of the random Turán theorem stated above and a sparse hypergraph removal lemma. Many similar results have recently been obtained independently in a different way by Schacht and by Friedgut, Rödl and Schacht.

1   Introduction

In recent years there has been a trend in combinatorics towards proving that certain well-known theorems, such as Ramsey's theorem, Turán's theorem and Szemerédi's theorem, have "sparse random" analogues. For instance, the first non-trivial case of Turán's theorem asserts that a subgraph of K_n with more than $\lfloor n/2 \rfloor \lceil n/2 \rceil$ edges must contain a triangle. A sparse random analogue of this theorem is the assertion that if one defines a random subgraph G of K_n by choosing each edge independently at random with some very small probability p, then with high probability every subgraph H of G such that $|E(H)| \geq \left(\frac{1}{2} + \epsilon\right)|E(G)|$ will contain a triangle. Several results of this kind have been proved, and in some cases, including this one, the exact bounds on what p one can take are known up to a constant factor.

The greatest success in this line of research has been with analogues of Ramsey's theorem [43]. Recall that Ramsey's theorem (in one of its many forms) states that, for every graph H and every natural number r, there exists n such that if the edges of the complete graph K_n are coloured with r colours, then there must be a copy of H with all its edges of the same colour. Such a copy of H is called *monochromatic*.

Let us say that a graph G is (H, r)-*Ramsey* if, however the edges of G are coloured with r colours, there must be a monochromatic copy of H. After efforts by several researchers [9, 39, 45, 46, 47], most notably Rödl and Ruciński, the following impressive theorem, a "sparse random version" of Ramsey's theorem, is now known. We write $G_{n,p}$ for the standard binomial model of random graphs, where each

---

edge in an n-vertex graph is chosen independently with probability p. We also write v_H and e_H for the number of vertices and edges, respectively, in a graph H.

Theorem 1.1. *Let $r \geq 2$ be a natural number and let H be a graph that is not a forest consisting of stars and paths of length 3. Then there exist positive constants c and C such that*

$$\lim_{n \to \infty} \mathbb{P}(G_{n,p} \text{ is } (H,r)\text{-Ramsey}) = \begin{cases} 0, & \text{if } p < cn^{-1/m_2(H)}, \\ 1, & \text{if } p > Cn^{-1/m_2(H)}, \end{cases}$$

where

$$m_2(H) = \max_{K \subset H, v_K \geq 3} \frac{e_K - 1}{v_K - 2}.$$

That is, given a graph G that is not a disjoint union of stars and paths of length 3, there is a threshold at approximately $p = n^{-1/m_2(H)}$ where the probability that the random graph $G_{n,p}$ is (H,r)-Ramsey changes from 0 to 1.

This theorem comes in two parts: the statement that above the threshold the graph is almost certainly (H,r)-Ramsey and the statement that below the threshold it almost certainly is not. We shall follow standard practice and call these the 1-*statement* and the 0-*statement*, respectively.

There have also been some efforts towards proving sparse random versions of Turán's theorem, but these have up to now been less successful. Turán's theorem [62], or rather its generalization, the Erdős-Stone-Simonovits theorem (see for example [3]), states that if H is some fixed graph, then any graph with n vertices that contains more than

$$\left(1 - \frac{1}{\chi(H) - 1} + o(1)\right) \binom{n}{2}$$

edges must contain a copy of H. Here, $\chi(H)$ is the chromatic number of H.

Let us say that a graph G is (H, ϵ)-*Turán* if every subgraph of G with at least

$$\left(1 - \frac{1}{\chi(H) - 1} + \epsilon\right) e(G)$$

edges contains a copy of H. One may then ask for the threshold at which a random graph becomes (H, ϵ)-Turán. The conjectured answer [26, 27, 36] is that the threshold is the same as it is for the corresponding Ramsey property.

Conjecture 1.2. *For every $\epsilon > 0$ and every graph H there exist positive constants c and C such that*

$$\lim_{n \to \infty} \mathbb{P}(G_{n,p} \text{ is } (H,\epsilon)\text{-Turán}) = \begin{cases} 0, & \text{if } p < cn^{-1/m_2(H)}, \\ 1, & \text{if } p > Cn^{-1/m_2(H)}, \end{cases}$$

where

$$m_2(H) = \max_{K \subset H, v_K \geq 3} \frac{e_K - 1}{v_K - 2}.$$

A difference between this conjecture and Theorem 1.1 is that the 0-statement in this conjecture is very simple to prove. To see this, suppose that p is such that the expected number of copies of H in $G_{n,p}$ is significantly less than the expected number of edges in $G_{n,p}$. Then, since the number of copies of H and the number of edges are both concentrated around their expectations, we can almost always

remove a small number of edges from $G_{n,p}$ and get rid of all copies of H, which proves that $G_{n,p}$ is not (H, ϵ)-Turán. The expected number of copies of H (if we label the vertices of H) is approximately $n^{v_H} p^{e_H}$, while the expected number of edges in $G_{n,p}$ is approximately pn^2. The former becomes less than the latter when $p = n^{-(v_H-2)/(e_H-1)}$.

A further observation raises this bound. Suppose, for example, that H is a triangle with an extra edge attached to one of its vertices. It is clear that the real obstacle to finding copies of H is finding triangles: it is not hard to add edges to them. More generally, if H has a subgraph K with $\frac{e_K-1}{v_K-2} > \frac{e_H-1}{v_H-2}$, then we can increase our estimate of p to $n^{-(v_K-2)/(e_K-1)}$, since if we can get rid of copies of K then we have got rid of copies of H. Beyond this extra observation, there is no obvious way of improving the bound for the 0-statement, which is why it is the conjectured upper bound as well.

An argument along these lines does not work at all for the Ramsey property, since if one removes a few edges in order to eliminate all copies of H in one colour, then one has to give them another colour. Since the set of removed edges is likely to look fairly random, it is not at all clear that this can be done in such a way as to eliminate all monochromatic copies of H.

Conjecture 1.2 is known to be true for some graphs, for example K_3, K_4, K_5 (see [9, 36, 18], respectively) and all cycles (see [13, 26, 27]), but it is open in general. Some partial results towards the general conjecture, where the 1-statement is proved with a weaker exponent, have been given by Kohayakawa, Rödl and Schacht [37] and Szabó and Vu [58]. The paper of Szabó and Vu contains the best known upper bound in the case where H is the complete graph K_t for some $t \geq 6$; the bound they obtain is $p = n^{-1/(t-1.5)}$, whereas the conjectured best possible bound is $p = n^{-2/(t+1)}$ (since $m_2(K_t)$ works out to be $(t+1)/2$). Thus, there is quite a significant gap. The full conjecture has also been proved to be a consequence of the so-called KŁR conjecture [36] of Kohayakawa, Łuczak and Rödl. This conjecture, regarding the number of H-free graphs of a certain type, remains open, except in a few special cases [16, 17, 18, 35].[1]

As noted in [34, 36], the KŁR conjecture would also imply the following structural result about H-free graphs which contain nearly the extremal number of edges. The analogous result in the dense case, due to Simonovits [57], is known as the stability theorem. Roughly speaking, it says that if an H-free graph contains almost $\left(1 - \frac{1}{\chi(H)-1}\right)\binom{n}{2}$ edges, then it must be very close to being $(\chi(H)-1)$-partite.

Conjecture 1.3. *Let H be a graph with $\chi(H) \geq 3$ and let*

$$m_2(H) = \max_{K \subset H, v_K \geq 3} \frac{e_K - 1}{v_K - 2}.$$

Then, for every $\delta > 0$, there exist positive constants ϵ and C such that if G is a random graph on n vertices, where each edge is chosen independently with probability p at least $Cn^{-1/m_2(H)}$, then, with probability tending to 1 as n tends to infinity, every H-free subgraph of G with at least $\left(1 - \frac{1}{\chi(H)-1} - \epsilon\right)e(G)$ edges may be made $(\chi(H)-1)$-partite by removing at most δpn^2 edges.

Another example where some success has been achieved is Szemerédi's theorem [59]. This celebrated theorem states that, for every positive real number δ and every natural number k, there exists

---

[1]The full KŁR conjecture was subsequently established by Balogh, Morris and Samotij [1] and by Saxton and Thomason [54] (see also [6]). Their methods also allow one to give alternative proofs for many of the results in this paper. We refer the reader to [5] for a more complete overview.

a positive integer n such that every subset of the set $[n] = \{1, 2, \cdots, n\}$ of size at least δn contains a k-term arithmetic progression. The particular case where $k = 3$ had been proved much earlier by Roth [52], and is accordingly known as Roth's theorem. A sparse random version of Roth's theorem was proved by Kohayakawa, Łuczak and Rödl [35]. To state the theorem, let us say that a subset I of the integers is δ-*Roth* if every subset of I of size $\delta|I|$ contains a 3-term arithmetic progression. We shall also write $[n]_p$ for a random set in which each element of $[n]$ is chosen independently with probability p.

Theorem 1.4. *For every $\delta > 0$ there exist positive constants c and C such that*

$$\lim_{n \to \infty} \mathbb{P}([n]_p \ is \ \delta\text{-}Roth) = \begin{cases} 0, & if \ p < cn^{-1/2}, \\ 1, & if \ p > Cn^{-1/2}. \end{cases}$$

Once again the 0-statement is trivial (as it tends to be for density theorems): if $p = n^{-1/2}/2$, then the expected number of 3-term progressions in $[n]_p$ is less than $n^{1/2}/8$, while the expected number of elements of $[n]_p$ is $n^{1/2}/2$. Therefore, one can almost always remove an element from each progression and still be left with at least half the elements of $[n]_p$.

For longer progressions, the situation has been much less satisfactory. Let us define a set I of integers to be (δ, k)-*Szemerédi* if every subset of I of cardinality at least $\delta|I|$ contains a k-term arithmetic progression. Until recently, hardly anything was known at all about which random sets were (δ, k)-Szemerédi. However, that changed with the seminal paper of Green and Tao [24], who, on the way to proving that the primes contain arbitrarily long arithmetic progressions, showed that every *pseudo*random set is (δ, k)-Szemerédi, if "pseudorandom" is defined in an appropriate way. Their definition of pseudorandomness is somewhat complicated, but it is straightforward to show that quite sparse random sets are pseudorandom in their sense. From this the following result follows, though we are not sure whether it has appeared explicitly in print.

Theorem 1.5. *For every $\delta > 0$ and every $k \in \mathbb{N}$ there exists a function $p = p(n)$ tending to zero with n such that*

$$\lim_{n \to \infty} \mathbb{P}([n]_p \ is \ (\delta, k)\text{-}Szemerédi) = 1.$$

The approach of Green and Tao depends heavily on the use of a set of norms known as *uniformity norms*, introduced in [19]. In order to deal with k-term arithmetic progressions, one must use a uniformity norm that is based on a count of certain configurations that can be thought of as $(k-1)$-dimensional parallelepipeds. These configurations have k degrees of freedom (one for each dimension and one because the parallelepipeds can be translated) and size 2^{k-1}. A simple argument (similar to the arguments for the 0-statements in the density theorems above) shows that the best bound that one can hope to obtain by their methods is therefore at most $p = n^{-k/2^{k-1}}$. This is far larger than the bound that arises in the obvious 0-statement for Szemerédi's theorem: the same argument that gives a bound of $cn^{-1/2}$ for the Roth property gives a bound of $cn^{-1/(k-1)}$ for the Szemerédi property. However, even $p = n^{-k/2^{k-1}}$ is not the bound that they actually obtain, because they need in addition a "correlation condition" that is not guaranteed by the smallness of the uniformity norm. This means that the bound they obtain is of the form $n^{-o(1)}$.

The natural conjecture is that the obvious bound for the 0-statement is in fact correct, so it is far stronger than the bound of Green and Tao.

Conjecture 1.6. *For every $\delta > 0$ and every positive integer $k \geq 3$, there exist positive constants c and C such that*

$$\lim_{n \to \infty} \mathbb{P}([n]_p \text{ is } (\delta, k)\text{-Szemerédi}) = \begin{cases} 0, & \text{if } p < cn^{-1/(k-1)}, \\ 1, & \text{if } p > Cn^{-1/(k-1)}. \end{cases}$$

One approach to proving Szemerédi's theorem is known as the hypergraph removal lemma. Proved independently by Nagle, Rödl, Schacht and Skokan [40, 51] and by the second author [21] (see also [61]), this theorem states that for every $\delta > 0$ and every positive integer $k \geq 2$ there exists a constant $\epsilon > 0$ such that if G is a k-uniform hypergraph containing at most ϵn^{k+1} copies of the complete k-uniform hypergraph $K_{k+1}^{(k)}$ on $k+1$ vertices, then it may be made $K_{k+1}^{(k)}$-free by removing at most δn^k edges. Once this theorem is known, Szemerédi's theorem follows as an easy consequence. The question of whether an analogous result holds within random hypergraphs was posed by Łuczak [38]. For $k = 2$, this follows from the work of Kohayakawa, Łuczak and Rödl [35].

Conjecture 1.7. *For every $\delta > 0$ and every integer $k \geq 2$ there exist constants $\epsilon > 0$ and C such that, if H is a random k-uniform hypergraph on n vertices where each edge is chosen independently with probability p at least $Cn^{-1/k}$, then, with probability tending to 1 as n tends to infinity, every subgraph of H containing at most $\epsilon p^{k+1} n^{k+1}$ copies of the complete k-uniform hypergraph $K_{k+1}^{(k)}$ on $k+1$ vertices may be made $K_{k+1}^{(k)}$-free by removing at most $\delta p n^k$ edges.*

1.1 The main results of this paper

In the next few sections we shall give a very general method for proving sparse random versions of combinatorial theorems. This method allows one to obtain sharp bounds for several theorems, of which the principal (but by no means only) examples are positive answers to the conjectures we have just mentioned. This statement comes with one caveat. When dealing with graphs and hypergraphs, we shall restrict our attention to those which are well-balanced in the following sense. Note that most graphs of interest, including complete graphs and cycles, satisfy this condition.

Definition 1.8. *A k-uniform hypergraph K is said to be strictly k-balanced if, for every subgraph L of K,*

$$\frac{e_K - 1}{v_K - k} > \frac{e_L - 1}{v_L - k}.$$

The main results we shall prove in this paper (in the order in which we discussed them above, but not the order in which we shall prove them) are as follows. The first is a sparse random version of Ramsey's theorem. Of course, as we have already mentioned, this is known: however, our theorem applies not just to graphs but to hypergraphs, where the problem was wide open apart from a few special cases [49, 50]. As we shall see, our methods apply just as easily to hypergraphs as they do to graphs. We write $G_{n,p}^{(k)}$ for a random k-uniform hypergraph on n vertices, where each hyperedge is chosen independently with probability p. If K is some fixed k-uniform hypergraph, we say that a hypergraph is (K, r)-Ramsey if every r-colouring of its edges contains a monochromatic copy of K.

Theorem 1.9. *Given a natural number r and a strictly k-balanced k-uniform hypergraph K, there exists a positive constant C such that*

$$\lim_{n \to \infty} \mathbb{P}(G_{n,p}^{(k)} \text{ is } (K, r)\text{-Ramsey}) = 1, \text{ if } p > Cn^{-1/m_k(K)},$$

where $m_k(K) = (e_K - 1)/(v_K - k)$.

One problem that the results of this paper leave open is to establish a corresponding 0-statement for Theorem 1.9. The above bound is the threshold below which the number of copies of K becomes less than the number of hyperedges, so the results for graphs make it highly plausible that the 0-statement holds when $p < cn^{-1/m_k(K)}$ for small enough c. However, the example of stars, for which the threshold is lower than expected, shows that we cannot take this result for granted.

We shall also prove Conjecture 1.2 for strictly 2-balanced graphs. In particular, it holds for complete graphs.

Theorem 1.10. *Given $\epsilon > 0$ and a strictly 2-balanced graph H, there exists a positive constant C such that*
$$\lim_{n \to \infty} \mathbb{P}(G_{n,p} \text{ is } (H, \epsilon)\text{-Turán}) = 1, \ \ \text{if } p > Cn^{-1/m_2(H)},$$
where $m_2(H) = (e_H - 1)/(v_H - 2)$.

A slightly more careful application of our methods also allows us to prove its structural counterpart, Conjecture 1.3, for strictly 2-balanced graphs.

Theorem 1.11. *Given a strictly 2-balanced graph H with $\chi(H) \geq 3$ and a constant $\delta > 0$, there exist positive constants C and ϵ such that in the random graph $G_{n,p}$ chosen with probability $p \geq Cn^{-1/m_2(H)}$, where $m_2(H) = (e_H-1)/(v_H-2)$, the following holds with probability tending to 1 as n tends to infinity. Every H-free subgraph of $G_{n,p}$ with at least $\left(1 - \frac{1}{\chi(H)-1} - \epsilon\right) e(G)$ edges may be made $(\chi(H) - 1)$-partite by removing at most δpn^2 edges.*

We also prove Conjecture 1.6, obtaining bounds for the Szemerédi property that are essentially best possible.

Theorem 1.12. *Given $\delta > 0$ and a natural number $k \geq 3$, there exists a constant C such that*
$$\lim_{n \to \infty} \mathbb{P}([n]_p \text{ is } (\delta, k)\text{-Szemerédi}) = 1, \ \ \text{if } p > Cn^{-1/(k-1)}.$$

Our final main result is a proof of Conjecture 1.7, the sparse hypergraph removal lemma. As we have mentioned, the dense hypergraph removal lemma implies Szemerédi's theorem, but it turns out that the sparse hypergraph removal lemma does not imply Theorem 1.12. The difficulty is this. When we prove Szemerédi's theorem using the removal lemma, we first pass to a hypergraph to which the removal lemma can be applied. Unfortunately, in the sparse case, passing from the sparse random set to the corresponding hypergraph gives us a sparse hypergraph with dependencies between its edges, whereas in the sparse hypergraph removal lemma we assume that the edges of the sparse random hypergraph are independent. While it is likely that this problem can be overcome, we did not, in the light of Theorem 1.12, see a strong reason for doing so.

In addition to these main results, we shall discuss other density theorems, such as Turán's theorem for hypergraphs (where, even though the correct bounds are not known in the dense case, we can obtain the threshold at which the bounds in the sparse random case will be the same), the multidimensional Szemerédi theorem of Furstenberg and Katznelson [15] and the Bergelson-Leibman theorem [2] concerning polynomial configurations in dense sets. In the colouring case, we shall discuss Schur's theorem [56] as a further example. Note that many similar results have also been obtained by a different method by Schacht [55] and by Friedgut, Rödl and Schacht [12].

1.2 A preliminary description of the argument

The basic idea behind our proof is to use a *transference principle* to deduce sparse random versions of density and colouring results from their dense counterparts. To oversimplify slightly, a transference principle in this context is a statement along the following lines. Let X be a structure such as the complete graph K_n or the set $\{1, 2, \ldots, n\}$, and let U be a sparse random subset of X. Then, for every subset $A \subset U$, there is a subset $B \subset X$ that has similar properties to A. In particular, the density of B is approximately the same as the relative density of A in U, and the number of substructures of a given kind in A is an appropriate multiple of the number of substructures of the same kind in B.

Given a strong enough principle of this kind, one can prove a sparse random version of Szemerédi's theorem, say, as follows. Let A be a subset of $[n]_p$ of relative density δ. Then there exists a subset B of $[n]$ of size approximately δn such that the number of k-term progressions in B is approximately p^{-k} times the number of k-term progressions in A. From Szemerédi's theorem it can be deduced that the number of k-term progressions in B is at least $c(\delta)n^2$, so the number of k-term progressions in A is at least $c(\delta)p^k n^2/2$. Since the size of A is about pn, we have roughly pn degenerate progressions. Hence, there are non-degenerate progressions within A as long as $p^k n^2$ is significantly larger than pn, that is, as long as p is at least $Cn^{-1/(k-1)}$ for some large C.

It is very important to the success of the above argument that a dense subset of $[n]$ should contain not just one progression but several, where "several" means a number that is within a constant of the trivial upper bound of n^2. The other combinatorial theorems discussed above have similarly "robust" versions and again these are essential to us. Very roughly, our general theorems say that a typical combinatorial theorem that is robust in this sense will have a sparse random version with an upper bound for the probability threshold that is very close to a natural lower bound that is trivial for density theorems and often true, even if no longer trivial, for Ramsey theorems.

It is also very helpful to have a certain degree of homogeneity. For instance, in order to prove the sparse version of Szemerédi's theorem we use the fact that it is equivalent to the sparse version of Szemerédi's theorem in $\mathbb{Z}_n$, where we have the nice property that for every k and every j with $1 \le j \le k$, every element x appears in the jth place of a k-term arithmetic progression in exactly n ways (or $n-1$ if you discount the degenerate progression with common difference 0). It will also be convenient to assume that n is prime, since in this case we know that for every pair of points x, y in $\mathbb{Z}_n$ there is exactly one arithmetic progression of length k that starts with x and ends in y. This simple homogeneity property will prove useful when we come to do our probabilistic estimates.

The idea of using a transference principle to obtain sparse random versions of robust combinatorial statements is not what is new about this paper. In fact, this was exactly the strategy of Green and Tao in their paper on the primes, and could be said to be the main idea behind their proof (though of course it took many further ideas to get it to work). Since it is difficult to say what is new about our argument without going into slightly more detail, we postpone further discussion for now. However, there are three further main ideas involved and we shall highlight them as they appear.

In the next few sections, we shall find a very general set of criteria under which one may transfer combinatorial statements to the sparse random setting. In Sections 5-8, we shall show how to prove that these criteria hold. Section 9 is a brief summary of the general results, both conditional and unconditional, that have been proved up to that point. In Section 10, we show how these results may be applied to prove the various theorems promised in the introduction. In Section 11, we conclude by briefly mentioning some questions that are still open.

1.3 Notation

We finish this section with some notation and terminology that we shall need throughout the course of the paper. By a *measure* on a finite set X we shall mean a non-negative function from X to $\mathbb{R}$. Usually our measures will have average value 1, or very close to 1. The *characteristic measure* μ of a subset U of X will be the function defined by $\mu(x) = |X|/|U|$ if $x \in U$ and $\mu(x) = 0$ otherwise.

Often our set U will be a random subset of X with each element of X chosen with probability p, the choices being independent. In this case, we shall use the shorthand $U = X_p$, just as we wrote $[n]_p$ for a random subset of $[n]$ in the statement of the sparse random version of Szemerédi's theorem earlier. When $U = X_p$ it is more convenient to consider the measure μ that is equal to p^{-1} times the characteristic function of U. That is, $\mu(x) = p^{-1}$ if $x \in U$ and 0 otherwise. To avoid confusion, we shall call this the *associated measure* of U. Strictly speaking, we should not say this, since it depends not just on U but on the value of p used when U was chosen, but this will always be clear from the context so we shall not bother to call it the associated measure of (U, p).

For an arbitrary function f from X to $\mathbb{R}$ we shall write $\mathbb{E}_x f(x)$ for $|X|^{-1} \sum_{x \in X} f(x)$. Note that if μ is the characteristic measure of a set U, then $\mathbb{E}_x \mu(x) = 1$ and $\mathbb{E}_x \mu(x) f(x) = \mathbb{E}_{x \in U} f(x)$ for any function f. If $U = X_p$ and μ is the associated measure of U, then we can no longer say this. However, we can say that the *expectation* of $\mathbb{E}_x \mu(x)$ is 1. Also, with very high probability the cardinality of U is roughly $p|X|$, so with high probability $\mathbb{E}_x \mu(x)$ is close to 1. More generally, if $|f(x)| \leq 1$ for every $x \in X$, then with high probability $\mathbb{E}_x \mu(x) f(x)$ is close to $\mathbb{E}_{x \in U} f(x)$. We also take expectations over several variables: if it is clear from the context that k variables $x_1, \ldots, x_k$ range over finite sets $X_1, \ldots, X_k$, respectively, then $\mathbb{E}_{x_1, \ldots, x_k}$ will be shorthand for $|X_1|^{-1} \ldots |X_k|^{-1} \sum_{x_1 \in X_1} \cdots \sum_{x_k \in X_k}$. If the range of a variable is not clear from the context then we shall specify it.

We define an inner product for real-valued functions on X by the formula $\langle f, g \rangle = \mathbb{E}_x f(x) g(x)$, and we define the L_p norm by $\|f\|_p = (\mathbb{E}_x |f(x)|^p)^{1/p}$. In particular, $\|f\|_1 = \mathbb{E}_x |f(x)|$ and $\|f\|_\infty = \max_x |f(x)|$.

Let $\|.\|$ be a norm on the space $\mathbb{R}^X$. The *dual norm* $\|.\|^*$ of $\|.\|$ is a norm on the collection of linear functionals ϕ acting on $\mathbb{R}^X$ given by

$$\|\phi\|^* = \sup\{|\langle f, \phi \rangle| : \|f\| \leq 1\}.$$

It follows trivially from this definition that $|\langle f, \phi \rangle| \leq \|f\| \|\phi\|^*$. Almost as trivially, it follows that if $|\langle f, \phi \rangle| \leq 1$ whenever $\|f\| \leq \eta$, then $\|\phi\|^* \leq \eta^{-1}$, a fact that will be used repeatedly.

2 Transference principles

As we have already mentioned, a central notion in this paper is that of transference. Roughly speaking, a *transference principle* is a theorem that states that every function f in one class can be replaced by a function g in another, more convenient class in such a way that the properties of f and g are similar.

To understand this concept and why it is useful, let us look at the sparse random version of Szemerédi's theorem that we shall prove. Instead of attacking this directly, it is convenient to prove a functional generalization of it. The statement we shall prove is the following.

Theorem 2.1. *For every positive integer k and every $\delta > 0$ there are positive constants c and C with the following property. Let $p \geq Cn^{-1/(k-1)}$, let U be a random subset of $\mathbb{Z}_n$ where each element is*

chosen independently with probability p and let μ be the associated measure of U. Then, with probability tending to 1 as n tends to infinity, every function f such that $0 \leq f \leq \mu$ and $\mathbb{E}_x f(x) \geq \delta$ satisfies the inequality

$$\mathbb{E}_{x,d} f(x) f(x+d) \ldots f(x+(k-1)d) \geq c.$$

To understand the normalization, it is a good exercise (and an easy one) to check that with high probability $\mathbb{E}_{x,d}\mu(x)\mu(x+d)\ldots\mu(x+(k-1)d)$ is close to 1, so that the conclusion of Theorem 2.1 is stating that $\mathbb{E}_{x,d}f(x)f(x+d)\ldots f(x+(k-1)d)$ is within a constant of its trivial maximum. (If p is smaller than $n^{-1/(k-1)}$ then this is no longer true: the main contribution to $\mathbb{E}_{x,d}\mu(x)\mu(x+d)\ldots\mu(x+(k-1)d)$ comes from the degenerate progressions where $d=0$.)

Our strategy for proving this theorem is to "transfer" the function f from the sparse set U to $\mathbb{Z}_n$ itself and then to deduce the conclusion from the following robust functional version of Szemerédi's theorem, which can be proved by a simple averaging argument due essentially to Varnavides [64].

Theorem 2.2. *For every $\delta > 0$ and every positive integer k there is a constant $c > 0$ such that, for every positive integer n, every function $g : \mathbb{Z}_n \to [0,1]$ with $\mathbb{E}_x g(x) \geq \delta$ satisfies the inequality*

$$\mathbb{E}_{x,d} g(x) g(x+d) \ldots g(x+(k-1)d) \geq c.$$

Note that in this statement we are no longer talking about dense subsets of $\mathbb{Z}_n$, but rather about $[0,1]$-valued functions defined on $\mathbb{Z}_n$ with positive expectation. It will be important in what follows that any particular theorem we wish to transfer has such an equivalent functional formulation. As we shall see in Section 4, all of the theorems that we consider do have such formulations.

Returning to transference principles, our aim is to find a function g with $0 \leq g \leq 1$ for which we can prove that $\mathbb{E}_x g(x) \approx \mathbb{E}_x f(x)$ and that

$$\mathbb{E}_{x,d} g(x) g(x+d) \ldots g(x+(k-1)d) \approx \mathbb{E}_{x,d} f(x) f(x+d) \ldots f(x+(k-1)d).$$

We can then argue as follows: if $\mathbb{E}_x f(x) \geq \delta$, then $\mathbb{E}_x g(x) \geq \delta/2$; by Theorem 2.2 it follows that $\mathbb{E}_{x,d} g(x) g(x+d) \ldots g(x+(k-1)d)$ is bounded below by a constant c; and this implies that $\mathbb{E}_{x,d} f(x) f(x+d) \ldots f(x+(k-1)d) \geq c/2$.

In the rest of this section we shall show how the Hahn-Banach theorem can be used to prove general transference principles. This was first demonstrated by the second author in [22], and independently (in a slightly different language) by Reingold, Trevisan, Tulsiani and Vadhan [44], and leads to simpler proofs than the method used by Green and Tao. The first transference principle we shall prove is particularly appropriate for density theorems: this one was shown in [22] but for convenience we repeat the proof. Then we shall prove a modification of it for use with colouring theorems.

Let us begin by stating the finite-dimensional Hahn-Banach theorem in its separation version.

Lemma 2.3. *Let K be a closed convex set in $\mathbb{R}^n$ and let v be a vector that does not belong to K. Then there is a real number t and a linear functional ϕ such that $\phi(v) > t$ and such that $\phi(w) \leq t$ for every $w \in K$.*

The reason the Hahn-Banach theorem is useful to us is that one often wishes to prove that one function is a sum of others with certain properties, and often the sets of functions that satisfy those properties are convex (or can easily be made convex). For instance, we shall want to write a function f with $0 \leq f \leq \mu$ as a sum $g + h$ with $0 \leq g \leq 1$ and with h small in a certain norm. The following

lemma, an almost immediate consequence of Lemma 2.3, tells us what happens when a function *cannot* be decomposed in this way. We implicitly use the fact that every linear functional on $\mathbb{R}^Y$ has the form $f \mapsto \langle f, \phi \rangle$ for some ϕ.

Lemma 2.4. *Let Y be a finite set and let K and L be two subsets of $\mathbb{R}^Y$ that are closed and convex and that contain 0. Suppose that $f \notin K + L$. Then there exists a function $\phi \in \mathbb{R}^Y$ such that $\langle f, \phi \rangle > 1$ and such that $\langle g, \phi \rangle \leq 1$ for every $g \in K$ and $\langle h, \phi \rangle \leq 1$ for every $h \in L$.*

Proof. By Lemma 2.3 there is a function ϕ and a real number t such that $\langle f, \phi \rangle > t$ and such that $\langle g + h, \phi \rangle \leq t$ whenever $g \in K$ and $h \in L$. Setting $h = 0$ we deduce that $\langle g, \phi \rangle \leq t$ for every $g \in K$, and setting $g = 0$ we deduce that $\langle h, \phi \rangle \leq t$ for every $h \in L$. Setting $g = h = 0$ we deduce that $t \geq 0$. Dividing through by t (or by $\frac{1}{2}\langle f, \phi \rangle$ if $t = 0$) we see that we may take t to be 1. $\qquad \square$

Now let us prove our two transference principles, beginning with the density one. In the statement of the theorem below we write ϕ_+ for the positive part of ϕ.

Lemma 2.5. *Let ϵ and η be positive real numbers, let μ and ν be non-negative functions defined on a finite set X and let $\|.\|$ be a norm on $\mathbb{R}^X$. Suppose that $\langle \mu - \nu, \phi_+ \rangle \leq \epsilon$ whenever $\|\phi\|^* \leq \eta^{-1}$. Then for every function f with $0 \leq f \leq \mu$ there exists a function g with $0 \leq g \leq \nu$ such that $\|(1+\epsilon)^{-1}f - g\| \leq \eta$.*

Proof. If we cannot approximate $(1 + \epsilon)^{-1}f$ in this way, then we cannot write $(1 + \epsilon)^{-1}f$ as a sum $g + h$ with $0 \leq g \leq \nu$ and $\|h\| \leq \eta$. Now the sets $K = \{g : 0 \leq g \leq \nu\}$ and $L = \{h : \|h\| \leq \eta\}$ are closed and convex and they both contain 0. It follows from Lemma 2.4, with $Y = X$, that there is a function ϕ with the following three properties.

- $\langle (1 + \epsilon)^{-1}f, \phi \rangle > 1$;

- $\langle g, \phi \rangle \leq 1$ whenever $0 \leq g \leq \nu$;

- $\langle h, \phi \rangle \leq 1$ whenever $\|h\| \leq \eta$.

From the first of these properties we deduce that $\langle f, \phi \rangle > 1 + \epsilon$. From the second we deduce that $\langle \nu, \phi_+ \rangle \leq 1$, since the function g that takes the value $\nu(x)$ when $\phi(x) \geq 0$ and 0 otherwise maximizes the value of $\langle g, \phi \rangle$ over all $g \in K$. And from the third property we deduce immediately that $\|\phi\|^* \leq \eta^{-1}$.

But our hypothesis implies that $\langle \mu, \phi_+ \rangle \leq \langle \nu, \phi_+ \rangle + \epsilon$. It therefore follows that

$$1 + \epsilon < \langle f, \phi \rangle \leq \langle f, \phi_+ \rangle \leq \langle \mu, \phi_+ \rangle \leq \langle \nu, \phi_+ \rangle + \epsilon \leq 1 + \epsilon,$$

which is a contradiction. $\qquad \square$

Later we shall apply Lemma 2.5 with μ the associated measure of a sparse random set and ν the constant measure 1.

The next transference principle is the one that we shall use for obtaining sparse random colouring theorems. It may seem strange that the condition we obtain on $g_1 + \cdots + g_r$ is merely that it is less than ν (rather than equal to ν). However, we also show that f_i and g_i are close in a certain sense, and in applications that will imply that $g_1 + \cdots + g_r$ is indeed approximately equal to ν (which will be the constant measure 1). With a bit more effort, one could obtain equality from the Hahn-Banach method, but this would not make life easier later, since the robust versions of Ramsey theorems hold just as well when you colour almost everything as they do when you colour everything.

Lemma 2.6. *Let ϵ and η be positive real numbers, let r be a positive integer, let μ and ν be non-negative functions defined on a finite set X and let $\|.\|$ be a norm on $\mathbb{R}^X$. Suppose that $\langle \mu - \nu, (\max_{1 \leq i \leq r} \phi_i)_+ \rangle \leq \epsilon$ whenever $\phi_1, \ldots, \phi_r$ are functions with $\|\phi_i\|^* \leq \eta^{-1}$ for each i. Then for every sequence of r functions $f_1, \ldots, f_r$ with $f_i \geq 0$ for each i and $f_1 + \cdots + f_r \leq \mu$ there exist functions $g_1, \ldots, g_r$ with $g_i \geq 0$ for each i and $g_1 + \cdots + g_r \leq \nu$ such that $\|(1 + \epsilon)^{-1} f_i - g_i\| \leq \eta$ for each i.*

Proof. Suppose that the result does not hold for the r-tuple $(f_1, \ldots, f_r)$. Let K be the closed convex set of all r-tuples of functions $(g_1, \ldots, g_r)$ such that $g_i \geq 0$ for each i and $g_1 + \cdots + g_r \leq \nu$, and let L be the closed convex set of all r-tuples $(h_1, \ldots, h_r)$ such that $\|h_i\| \leq \eta$ for each i. Then both K and L contain 0 and our hypothesis is that $(1 + \epsilon)^{-1}(f_1, \ldots, f_r) \notin K + L$. Therefore, Lemma 2.4, with $Y = X^r$, gives us an r-tuple of functions $(\phi_1, \ldots, \phi_r)$ with the following three properties.

- $\sum_{i=1}^{r} \langle (1 + \epsilon)^{-1} f_i, \phi_i \rangle > 1$;

- $\sum_{i=1}^{r} \langle g_i, \phi_i \rangle \leq 1$ whenever $g_i \geq 0$ for each i and $g_1 + \cdots + g_r \leq \nu$;

- $\sum_{i=1}^{r} \langle h_i, \phi_i \rangle \leq 1$ whenever $\|h_i\| \leq \eta$ for each i.

The first of these conditions implies that $\sum_{i=1}^{r} \langle f_i, \phi_i \rangle > 1 + \epsilon$. In the second condition, let us choose the functions g_i as follows. For each x, pick an i such that $\phi_i(x)$ is maximal. If $\phi_i(x) \geq 0$, then set $g_i(x)$ to be $\nu(x)$, and otherwise set $g_i(x) = 0$. For each $j \neq i$, set $g_j(x)$ to be zero. Then $\sum_{i=1}^{r} g_i(x)\phi_i(x)$ is equal to $\nu(x) \max_i \phi_i(x)$ if this maximum is non-negative, and 0 otherwise. Therefore, $\sum_{i=1}^{r} \langle g_i, \phi_i \rangle = \langle \nu, (\max_i \phi_i)_+ \rangle$. Thus, it follows from the second condition that $\langle \nu, (\max_i \phi_i)_+ \rangle \leq 1$. Let us write ϕ for $\max_i \phi_i$. The third condition implies that $\|\phi_i\|^* \leq \eta^{-1}$ for each i.

Using this information together with our hypothesis about $\mu - \nu$, we find that

$$1 + \epsilon < \sum_{i=1}^{r} \langle f_i, \phi_i \rangle \leq \sum_{i=1}^{r} \langle f_i, \phi_+ \rangle \leq \langle \mu, \phi_+ \rangle \leq \langle \nu, \phi_+ \rangle + \epsilon \leq 1 + \epsilon,$$

a contradiction. □

3   The counting lemma

We now come to the second main idea of the paper, and perhaps the main new idea. Lemmas 2.5 and 2.6 will be very useful to us, but as they stand they are rather abstract: in order to make use of them we need to find a norm $\|.\|$ such that if $\|f - g\|$ is small then f and g behave similarly in a relevant way. Several norms have been devised for exactly this purpose, such as the uniformity norms mentioned earlier, and also "box norms" for multidimensional structures and "octahedral norms" for graphs and hypergraphs. It might therefore seem natural to try to apply Lemmas 2.5 and 2.6 to these norms. However, as we have already commented in the case of uniformity norms, if we do this then we cannot obtain sharp bounds: except in a few cases, these norms are related to counts of configurations that are too large to appear non-degenerately in very sparse random sets.

We are therefore forced to adopt a different approach. Instead of trying to use an off-the-shelf norm, we use a bespoke norm, designed to fit perfectly the problem at hand. Notice that Lemmas 2.5 and 2.6 become harder to apply as the norm $\|.\|$ gets bigger, since then the dual norm $\|.\|^*$ gets smaller and there are more functions ϕ with $\|\phi\|^* \leq \eta^{-1}$, and therefore more functions of the form

ϕ_+ for which one must show that $\langle \mu - \nu, \phi_+ \rangle \leq \epsilon$ (and similarly for $(\max_{1 \leq i \leq r} \phi_i)_+$ with colouring problems). Therefore, we shall try to make our norm as small as possible, subject to the condition we need it to satisfy: that f and g behave similarly if $\|f - g\|$ is small.

Thus, our norm will be defined by means of a universal construction. As with other universal constructions, this makes the norm easy to define but hard to understand concretely. However, we can get away with surprisingly little understanding of its detailed behaviour, as will become clear later. An advantage of this abstract approach is that it has very little dependence on the particular problem that is being studied: it is for that reason that we have ended up with a very general result.

Before we define the norm, let us describe the general set-up that we shall analyse. We shall begin with a finite set X and a collection S of ordered subsets of X, each of size k. Thus, any element $s \in S$ may be expressed in the form $s = (s_1, \ldots, s_k)$.

Here are two examples. When we apply our results to Szemerédi's theorem, we shall take X to be $\mathbb{Z}_n$, and S to be the set of ordered k-tuples of the form $(x, x+d, \ldots, x+(k-1)d)$, and when we apply it to Ramsey's theorem or Turán's theorem for K_4, we shall take X to be the edge set of the complete graph K_n and S to be the set of ordered sextuples of pairs of the form $(x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4)$, where x_1, x_2, x_3 and x_4 are vertices of K_n. Depending on the particular circumstance, we shall choose whether to include or ignore degenerate configurations. For example, for Szemerédi's theorem, it is convenient to include the possibility that $d = 0$, but for theorems involving K_4, we restrict to configurations where x_1, x_2, x_3 and x_4 are all distinct. In practice, it makes little difference, since the number of degenerate configurations is never very large.

In both these two examples, the collection S of ordered subsets of X has some nice homogeneity properties, which we shall assume for our general result because it makes the proofs cleaner, even if one sometimes has to work a little to show that these properties may be assumed.

Definition 3.1. *Let S be a collection of ordered k-tuples $s = (s_1, \ldots, s_k)$ of elements of a finite set X, and let us write $S_j(x)$ for the set of all s in S such that $s_j = x$. We shall say that S is* homogeneous *if for each j the sets $S_j(x)$ all have the same size.*

We shall assume throughout that our sets of ordered k-tuples are homogeneous in this sense. Note that this assumption does not hold for arithmetic progressions of length k if we work in the set $[n]$ rather than the set $\mathbb{Z}_n$. However, sparse random Szemerédi for $\mathbb{Z}_n$ implies sparse random Szemerédi for $[n]$, so this does not bother us. Similar observations can be used to convert several other problems into equivalent ones for which the set S is homogeneous. Moreover, such observations will easily accommodate any further homogeneity assumptions that we have to introduce in later sections.

The functional version of a combinatorial theorem about the ordered sets in S will involve expressions such as

$$\mathbb{E}_{s \in S} f(s_1) \ldots f(s_k).$$

Thus, what we wish to do is define a norm $\|.\|$ with the property that

$$\mathbb{E}_{s \in S} f(s_1) \ldots f(s_k) - \mathbb{E}_{s \in S} g(s_1) \ldots g(s_k)$$

can be bounded above in terms of $\|f - g\|$ whenever $0 \leq f \leq \mu$ and $0 \leq g \leq \nu$. This is what we mean by saying that f and g should behave similarly when $\|f - g\|$ is small.

The feature of the problem that gives us a simple and natural norm is the k-linearity of the expression $\mathbb{E}_{s \in S} f(s_1) \ldots f(s_k)$, which allows us to write the above difference as

$$\sum_{j=1}^{k} \mathbb{E}_{s \in S} g(s_1) \ldots g(s_{j-1})(f-g)(s_j) f(s_{j+1}) \ldots f(s_k).$$

Because we are assuming that the sets $S_j(x)$ all have the same size, we can write any expression of the form $\mathbb{E}_{s \in S} h_1(s_1) \ldots h_k(s_k)$ as

$$\mathbb{E}_{x \in X} h_j(x) \mathbb{E}_{s \in S_j(x)} h_1(s_1) \ldots h_{j-1}(s_{j-1}) h_{j+1}(s_{j+1}) \ldots h_k(s_k).$$

It will be very convenient to introduce some terminology and notation for expressions of the kind that are beginning to appear.

Definition 3.2. *Let X be a finite set and let S be a homogeneous collection of ordered subsets of X, each of size k. Then, given k functions $h_1, \ldots, h_k$ from X to $\mathbb{R}$, their jth convolution is defined to be the function*

$$*_j(h_1, \ldots, h_k)(x) = \mathbb{E}_{s \in S_j(x)} h_1(s_1) \ldots h_{j-1}(s_{j-1}) h_{j+1}(s_{j+1}) \ldots h_k(s_k).$$

We call this a convolution because in the special case where S is the set of arithmetic progressions of length 3 in $\mathbb{Z}_N$, we obtain convolutions in the conventional sense. Using this notation and the observation made above, we can rewrite

$$\mathbb{E}_{s \in S} g(s_1) \ldots g(s_{j-1})(f-g)(s_j) f(s_{j+1}) \ldots f(s_k)$$

as $\langle f - g, *_j(g \ldots, g, f, \ldots, f) \rangle$ (where it is understood that there are $j-1$ occurrences of g and $k-j$ occurrences of f), and from that we obtain the identity

$$\mathbb{E}_{s \in S} f(s_1) \ldots f(s_k) - \mathbb{E}_{s \in S} g(s_1) \ldots g(s_k) = \sum_{j=1}^{k} \langle f - g, *_j(g \ldots, g, f, \ldots, f) \rangle.$$

This, together with the triangle inequality, gives us the following lemma.

Lemma 3.3. *Let X be a finite set and let S be a homogeneous collection of ordered subsets of X, each of size k. Let f and g be two functions defined on X. Then*

$$|\mathbb{E}_{s \in S} f(s_1) \ldots f(s_k) - \mathbb{E}_{s \in S} g(s_1) \ldots g(s_k)| \leq \sum_{j=1}^{k} |\langle f - g, *_j(g, \ldots, g, f, \ldots, f) \rangle|.$$

It follows that if $f - g$ has small inner product with all functions of the form $*_j(g \ldots, g, f, \ldots, f)$, then $\mathbb{E}_{s \in S} f(s_1) \ldots f(s_k)$ and $\mathbb{E}_{s \in S} g(s_1) \ldots g(s_k)$ are close. It is tempting, therefore, to define a norm $\|.\|$ by taking $\|h\|$ to be the maximum value of $|\langle h, \phi \rangle|$ over all functions ϕ of the form $*_j(g, \ldots, g, f, \ldots, f)$ for which $0 \leq g \leq \nu$ and $0 \leq f \leq \mu$. If we did that, then we would know that $|\mathbb{E}_{s \in S} f(s_1) \ldots f(s_k) - \mathbb{E}_{s \in S} g(s_1) \ldots g(s_k)|$ was small whenever $\|f - g\|$ was small, which is exactly the property we need our norm to have. Unfortunately, this definition leads to difficulties. To see why we need to look in more detail at the convolutions.

Any convolution $*_j(g, \ldots, g, f, \ldots, f)$ is bounded above by $*_j(\nu, \ldots, \nu, \mu, \ldots, \mu)$. For the sake of example, let us consider the case of Szemerédi's theorem. Taking $\nu = 1$, we see that the jth convolution is bounded above by the function

$$P_j(x) = \mathbb{E}_d \mu(x + d) \ldots \mu(x + (k - j)d).$$

Up to normalization, this counts the number of progressions of length $k - j + 1$ beginning at x. If $j > 1$, probabilistic estimates imply that, at the critical probability $p = Cn^{-1/(k-1)}$, P_j is, with high probability, L_∞-bounded (that is, the largest value of the function is bounded by some absolute constant). However, functions of the form $*_1(f, \ldots, f)$ with $0 \le f \le \mu$ are almost always unbounded. This makes it much more difficult to control their inner products with $\mu - 1$, and we need to do that if we wish to apply the abstract transference principle from the previous section.

For graphs, a similar problem arises. The jth convolution will count, up to normalization, the number of copies of some subgraph of the given graph H that are rooted on a particular edge. If we assume that the graph is balanced, as we are doing, then, at probability $p = Cn^{-1/m_2(H)}$, this count will be L_∞-bounded for any proper subgraph of H. However, for H itself, we do not have this luxury and the function $*_1(f, \ldots, f)$ is again likely to be unbounded.

If we were prepared to increase the density of the random set by a polylogarithmic factor, we could ensure that even $*_1(f, \ldots, f)$ was bounded and this problem would go away. Thus, a significant part of the complication of this paper is due to our wish to get a bound that is best possible up to a constant.

There are two natural ways of getting around the difficulty if we are not prepared to sacrifice a polylogarithmic factor. One is to try to exploit the fact that although $*_1(f, \ldots, f)$ is not bounded, it typically takes large values very infrequently, so it is "close to bounded" in a certain sense. The other is to replace $*_1(f, \ldots, f)$ by a modification of the function that has been truncated at a certain maximum. It seems likely that both approaches can be made to work: we have found it technically easier to go for the second. The relevant definition is as follows.

Definition 3.4. *Let X be a finite set and let S be a homogeneous collection of ordered subsets of X, each of size k. Then, given k non-negative functions $h_1, \ldots, h_k$ from X to $\mathbb{R}$, their jth capped convolution $\circ_j(h_1, \ldots, h_k)$ is defined by*

$$\circ_j(h_1, \ldots, h_k)(x) = \min\{*_j(h_1, \ldots, h_k)(x), 2\}.$$

Unlike with ordinary convolutions, there is no obvious way of controlling the difference between $\mathbb{E}_{s \in S} f(s_1) \ldots f(s_k)$ and $\mathbb{E}_{s \in S} g(s_1) \ldots g(s_k)$ in terms of the inner product between $f - g$ and suitably chosen capped convolutions. So instead we shall look at a quantity that is related in a different way to the number of substructures of the required type. Roughly speaking, this quantity counts the number of substructures, but does not count too many if they start from the same point.

A natural quantity that fits this description is $\langle f, \circ_1(f, f, \ldots, f) \rangle$, and this is indeed closely related to the quantity we shall actually consider. However, there is an additional complication, which is that it is very convenient to think of our random set U as a union of m random sets $U_1, \ldots, U_m$, and of a function defined on U as an average $m^{-1}(f_1 + \cdots + f_m)$ of functions with f_i defined on U_i. More precisely, we shall take m independent random sets $U_1, \ldots, U_m$, each distributed as X_p. (Recall that X_p stands for a random subset of X where the elements are chosen independently with probability p.) Writing $\mu_1, \ldots, \mu_m$ for their associated measures, for each i we shall take a function f_i such that $0 \le f_i \le \mu_i$. Our assertion will then be about the average $f = m^{-1}(f_1 + \cdots + f_m)$. Note that

$0 \le f \le \mu$, where $\mu = m^{-1}(\mu_1 + \cdots + \mu_m)$, and that every function f with $0 \le f \le \mu$ can be expressed as an average of functions f_i with $0 \le f_i \le \mu_i$. Note also that if $U = U_1 \cup \cdots \cup U_m$ then μ is neither the characteristic measure of U nor the associated measure of U. However, provided p is fairly small, it is close to both with high probability, and this is all that matters.

Having chosen f in this way, the quantity we shall then look at is

$$\langle f, m^{-(k-1)} \sum_{i_2,\dots,i_k} \circ_1(f_{i_2},\dots,f_{i_k}) \rangle = \mathbb{E}_{i_1,\dots,i_k \in \{1,\dots,m\}} \langle f_{i_1}, \circ_1(f_{i_2},\dots,f_{i_k}) \rangle.$$

In other words, we expand the expression $\langle f, *_1(f, f, \dots, f) \rangle$ in terms of $f_1, \dots, f_m$ and then do the capping term by term.

Central to our approach is a "counting lemma", which is an easy corollary of the following result, which keeps track of the errors that are introduced by our "capping". (To understand the statement, observe that if we replaced the capped convolutions $\circ_j$ by their "genuine" counterparts $*_j$, then the two quantities that we are comparing would become equal.) In the next lemma, we assume that a homogeneous set S of ordered k-tuples has been given.

Lemma 3.5. *Let $\eta > 0$, let $m \ge 2k^3/\eta$ and let $\mu_1, \dots, \mu_m$ be non-negative functions defined on X with $\|\mu_i\|_1 \le 2$ for all i. Suppose that $\| *_1 (\mu_{i_2}, \dots, \mu_{i_k}) - \circ_1(\mu_{i_2}, \dots, \mu_{i_k}) \|_1 \le \eta$ whenever $i_2, \dots, i_k$ are distinct integers between 1 and m, and also that $*_j(1, 1, \dots, 1, \mu_{i_{j+1}}, \dots, \mu_{i_k})$ is uniformly bounded above by 2 whenever $j \ge 2$ and $i_{j+1}, \dots, i_k$ are distinct. For each i, let f_i be a function with $0 \le f_i \le \mu_i$, let $f = \mathbb{E}_i f_i$ and let g be a function with $0 \le g \le 1$. Then*

$$\mathbb{E}_{i_1,\dots,i_k \in \{1,\dots,m\}} \langle f_{i_1}, \circ_1(f_{i_2},\dots,f_{i_k}) \rangle - \langle g, *_1(g,g,\dots,g) \rangle$$

differs from

$$\sum_{j=1}^{k} \langle f - g, \mathbb{E}_{i_{j+1},\dots,i_k} \circ_j (g,g,\dots,g,f_{i_{j+1}},\dots,f_{i_k}) \rangle$$

by at most 2η.

Proof. Note first that

$$\begin{aligned} \mathbb{E}_{i_1,\dots,i_k} \langle f_{i_1}, \circ_1(f_{i_2},\dots,f_{i_k}) \rangle &= \mathbb{E}_{i_1,\dots,i_k} \langle f_{i_1} - g, \circ_1(f_{i_2},\dots,f_{i_k}) \rangle + \mathbb{E}_{i_2,\dots,i_k} \langle g, \circ_1(f_{i_2},\dots,f_{i_k}) \rangle \\ &= \mathbb{E}_{i_2,\dots,i_k} \langle f - g, \circ_1(f_{i_2},\dots,f_{i_k}) \rangle + \mathbb{E}_{i_2,\dots,i_k} \langle g, \circ_1(f_{i_2},\dots,f_{i_k}) \rangle. \end{aligned}$$

Since $0 \le *_1(f_{i_2},\dots,f_{i_k}) \le *_1(\mu_{i_2},\dots,\mu_{i_k})$, our assumption implies that, whenever $i_2,\dots,i_k$ are distinct, $\| *_1 (f_{i_2},\dots,f_{i_k}) - \circ_1(f_{i_2},\dots,f_{i_k}) \|_1 \le \eta$. In this case, therefore,

$$0 \le \langle g, *_1(f_{i_2},\dots,f_{i_k}) \rangle - \langle g, \circ_1(f_{i_2},\dots,f_{i_k}) \rangle \le \eta.$$

We also know that $\langle g, *_1(f_{i_2},\dots,f_{i_k}) \rangle = \langle f_{i_2}, *_2(g, f_{i_3},\dots,f_{i_k}) \rangle$ and that if $i_3,\dots,i_k$ are distinct then $*_2(g, f_{i_3},\dots,f_{i_k}) = \circ_2(g, f_{i_3},\dots,f_{i_k})$. Therefore,

$$0 \le \langle f_{i_2}, \circ_2(g, f_{i_3},\dots,f_{i_k}) \rangle - \langle g, \circ_1(f_{i_2},\dots,f_{i_k}) \rangle \le \eta.$$

15

Now the assumption that $*_j(1, 1, \ldots, 1, \mu_{i_{j+1}}, \ldots, \mu_{i_k})$ is bounded above by 2 whenever $j \geq 2$ and $i_{j+1}, \ldots, i_k$ are distinct implies that $\circ_j(g, g, \ldots, g, f_{i_{j+1}}, \ldots, f_{i_k})$ and $*_j(g, g, \ldots, g, f_{i_{j+1}}, \ldots, f_{i_k})$ are equal under these circumstances. From this it is a small exercise to show that

$$\langle f_{i_2}, \circ_2(g, f_{i_3}, \ldots, f_{i_k}) \rangle - \langle g, \circ_k(g, g, \ldots, g) \rangle = \sum_{j=2}^{k} \langle f_{i_j} - g, \circ_j(g, g, \ldots, g, f_{i_{j+1}}, \ldots, f_{i_k}) \rangle.$$

Therefore, for $i_2, \ldots, i_k$ distinct,

$$\langle g, \circ_1(f_{i_2}, \ldots, f_{i_k}) \rangle - \langle g, \circ_k(g, g, \ldots, g) \rangle \tag{1}$$

differs from

$$\sum_{j=2}^{k} \langle f_{i_j} - g, \circ_j(g, g, \ldots, g, f_{i_{j+1}}, \ldots, f_{i_k}) \rangle \tag{2}$$

by at most η.

The probability that $i_1, \ldots, i_k$ are not distinct is at most $\binom{k}{2} m^{-1} \leq \eta/4k$, and if they are not distinct then the difference between (1) and (2) is certainly no more than $4k$ (since all capped convolutions take values in $[0, 2]$ and $\|f_{i_j}\|_1 \leq \|\mu_{i_j}\|_1 \leq 2$). Therefore, taking the expectation over all $(i_1, \ldots, i_k)$ (not necessarily distinct) and noting that $\langle g, \circ_k(g, g, \ldots, g) \rangle = \langle g, *_1(g, g, \ldots, g) \rangle$, we find that

$$\mathbb{E}_{i_1, \ldots, i_k} \langle f_{i_1}, \circ_1(f_{i_2}, \ldots, f_{i_k}) \rangle - \langle g, *_1(g, g, \ldots, g) \rangle$$

differs from

$$\sum_{j=1}^{k} \langle f - g, \mathbb{E}_{i_{j+1}, \ldots, i_k} \circ_j (g, g, \ldots, g, f_{i_{j+1}}, \ldots, f_{i_k}) \rangle$$

by at most 2η, as claimed. $\qquad\square$

To state our counting lemma, we need to define the norm that we shall actually use.

Definition 3.6. *Let X be a finite set and let S be a homogeneous collection of ordered subsets of X, each of size k. Let $\mu = (\mu_1, \ldots, \mu_m)$ be a sequence of measures on X. A $(\mu, 1)$-basic anti-uniform function is a function of the form $\circ_j(g, \ldots, g, f_{i_{j+1}}, \ldots, f_{i_k})$, where $1 \leq j \leq k$, $i_{j+1}, \ldots, i_k$ are distinct, $0 \leq g \leq 1$ and $0 \leq f_{i_h} \leq \mu_{i_h}$ for every h between $j + 1$ and k. Let $\Phi_{\mu,1}$ be the set of all $(\mu, 1)$-basic anti-uniform functions and define the norm $\|.\|_{\mu,1}$ by taking $\|h\|_{\mu,1}$ to be $\max\{|\langle h, \phi \rangle| : \phi \in \Phi_{\mu,1}\}$.*

The phrase "basic anti-uniform function" is borrowed from Green and Tao, since our basic anti-uniform functions are closely related to functions of the same name that appear in their paper [24].

Our counting lemma is now as follows. It says that if $\|f - g\|_{\mu,1}$ is small, then the "sparse" expression given by $\mathbb{E}_{i_1, \ldots, i_k \in \{1, \ldots, m\}} \langle f_{i_1}, \circ_1(f_{i_2}, \ldots, f_{i_k}) \rangle$ is approximated by the "dense" expression $\langle g, *_1(g, g, \ldots, g) \rangle$. This lemma modifies Lemma 3.3 in two ways: it splits f up into $m^{-1}(f_1 + \cdots + f_m)$ and it caps all the convolutions that appear when one expands out the expression $\langle f, *_1(f, \ldots, f) \rangle$ in terms of the f_i.

Corollary 3.7. *Suppose that the assumptions of Lemma 3.5 hold, and that $|\langle f - g, \phi \rangle| \leq \eta/k$ for every basic anti-uniform function $\phi \in \Phi_{\mu,1}$. Then $\mathbb{E}_x g(x) \geq \mathbb{E}_x f(x) - \eta/k$, and*

$$\left| \mathbb{E}_{i_1, \ldots, i_k \in \{1, \ldots, m\}} \langle f_{i_1}, \circ_1(f_{i_2}, \ldots, f_{i_k}) \rangle - \langle g, *_1(g, g, \ldots, g) \rangle \right| \leq 4\eta.$$

Proof. The function $\circ_k(1, 1, \ldots, 1)$ is a basic anti-uniform function, and it takes the constant value 1. Since $\mathbb{E}_x h(x) = \langle h, 1 \rangle$ for any function h, this implies the first assertion.

Now the probability that $i_1, \ldots, i_k$ are distinct is again at most $\eta/4k$, and if they are not distinct we at least know that $|\langle f - g, \circ_j(g, g, \ldots, g, f_{i_{j+1}}, \ldots, f_{i_k}) \rangle| \leq 4$. Therefore, our hypothesis also implies that

$$\sum_{j=1}^{k} |\langle f - g, \mathbb{E}_{i_{j+1}, \ldots, i_k} \circ_j (g, g, \ldots, g, f_{i_{j+1}}, \ldots, f_{i_k}) \rangle| \leq k(\eta/k) + 4k(\eta/4k) = 2\eta.$$

Combining this with Lemma 3.5, we obtain the result. □

In order to prove analogues of structural results such as the Erdős-Simonovits stability theorem and the hypergraph removal lemma we shall need to preserve slightly more information when we replace our sparsely supported function f by a densely supported function g. For example, to prove the stability theorem, we proceed as follows. Given a subgraph A of the random graph $G_{n,p}$, we create a weighted subgraph B of K_n that contains the same number of copies of H, up to normalization. However, to make the proof work, we also need the edge density of B within any large vertex set to correspond to the edge density of A within that set. Suppose that we have this property as well and that A is H-free. Then B has very few copies of H. A robust version of the stability theorem then tells us that B may be made $(\chi(H) - 1)$-partite by removing a small number of edges (or rather a small weight of weighted edges). Let us look at the resulting weighted graph B'. It consists of $\chi(H) - 1$ vertex sets, all of which have zero weight inside. Therefore, in B, each of these sets had only a small weight to begin with. Since all "local densities" of A reflect those of B, these vertex sets contain only a very small proportion of the edges in A as well. Removing these edges makes A into a $(\chi(H) - 1)$-partite graph and we are done.

How do we ensure that local densities are preserved? All we have to do is enrich our set of basic anti-uniform functions by adding an appropriate set of functions that will allow us to transfer local densities from the sparse structure to the dense one. For example, in the case above we need to know that A and B have roughly the same inner product (when appropriately weighted) with the characteristic function of the complete graph on any large set V of vertices. We therefore add these characteristic functions to our stock of basic anti-uniform functions. For other applications, we need to maintain more intricate local density conditions. However, as we shall see, as long as the corresponding set of additional functions is sufficiently small, this does not pose a problem.

4 A conditional proof of the main theorems

In this section, we shall collect together the results of Sections 2 and 3 in order to make clear what is left to prove. We start with a simple and general lemma about duality in normed spaces.

Lemma 4.1. *Let Φ be a bounded set of real-valued functions defined on a finite set X such that the linear span of Φ is $\mathbb{R}^X$. Let a norm on $\mathbb{R}^X$ be defined by $\|f\| = \max\{|\langle f, \phi \rangle| : \phi \in \Phi\}$. Let $\|.\|^*$ be the dual norm. Then $\|\psi\|^* \leq 1$ if and only if ψ belongs to the closed convex hull of $\Phi \cup (-\Phi)$.*

Proof. If $\psi = \sum_i \lambda_i \phi_i$ with $\phi_i \in \Phi \cup (-\Phi)$, $\lambda_i \geq 0$ for each i and $\sum_i \lambda_i = 1$, and if $\|f\| \leq 1$, then $|\langle f, \psi \rangle| \leq \sum_i \lambda_i |\langle f, \phi_i \rangle| \leq 1$. The same is then true if ψ belongs to the closure of the convex hull of $\Phi \cup (-\Phi)$.

If ψ does not belong to this closed convex hull, then by the Hahn-Banach theorem there must be a function f such that $|\langle f, \phi \rangle| \leq 1$ for every $\phi \in \Phi$ and $\langle f, \psi \rangle > 1$. The first condition tells us that $\|f\| \leq 1$, so the second implies that $\|\psi\|^* > 1$. $\qquad \square$

So we already know a great deal about functions ϕ with bounded dual norm. Recall, however, that we must consider positive parts of such functions: we would like to show that $\langle \mu - \nu, \phi_+ \rangle$ is small whenever $\|\phi\|^*$ is of reasonable size. We need the following extra lemma to gain some control over these.

Lemma 4.2. *Let Ψ be a set of functions that take values in $[-2,2]$ and let $\epsilon > 0$. Then there exist constants d and M, depending on ϵ only, such that for every function ψ in the convex hull of Ψ, there is a function ω that belongs to M times the convex hull of all products $\pm\phi_1 \ldots \phi_j$ with $j \leq d$ and $\phi_1, \ldots, \phi_j \in \Psi$, such that $\|\psi_+ - \omega\|_\infty < \epsilon$.*

Proof. We start with the well-known fact that continuous functions on closed bounded intervals can be uniformly approximated by polynomials. Therefore, if $K(x)$ is the function defined on $[-2,2]$ that takes the value 0 if $x \leq 0$ and x if $x \geq 0$, then there is a polynomial P such that $|P(x) - K(x)| \leq \epsilon$ for every $x \in [-2,2]$. It follows that if ψ is a function that takes values in $[-2,2]$, then $\|P(\psi) - \psi_+\|_\infty \leq \epsilon$.

Let us apply this observation in the case where ψ is a convex combination $\sum_i \lambda_i \phi_i$ of functions $\phi_i \in \Psi$. If $P(t) = \sum_{j=1}^d a_j t^j$, then

$$P(\psi) = \sum_{j=1}^d a_j \sum_{i_1, \ldots, i_j} \lambda_{i_1} \ldots \lambda_{i_j} \phi_{i_1} \ldots \phi_{i_j}.$$

But $\sum_{i_1, \ldots, i_j} \lambda_{i_1} \ldots \lambda_{i_j} = 1$ for every j, so this proves that we can take M to be $\sum_{j=1}^d |a_j|$. This bound and the degree d depend on ϵ only, as claimed. $\qquad \square$

Similarly, for colouring problems, where we need to deal with the function $(\max_{1 \leq i \leq r} \phi_i)_+$, we have the following lemma. The proof is very similar to that of Lemma 4.2, though we must replace the function $K(x)$ that has to be approximated with the function $K(x_1, \ldots, x_r) = \max\{0, x_1, \ldots, x_r\}$ and apply a multivariate version of the uniform approximation theorem inside the set $[-2,2]^r$ (though the case we actually need follows easily from the one-dimensional theorem).

Lemma 4.3. *Let Ψ be a set of functions that take values in $[-2,2]$ and let $\epsilon > 0$. Then there exist constants d and M, depending on ϵ only, such that for every set of functions $\psi_1, \ldots, \psi_r$ in the convex hull of Ψ, there is a function ω that belongs to M times the convex hull of all products $\pm\phi_1 \ldots \phi_j$ with $j \leq d$ and $\phi_1, \ldots, \phi_j \in \Psi$, such that $\|(\max_{1 \leq i \leq r} \psi_i)_+ - \omega\|_\infty < \epsilon$.* $\qquad \square$

We shall split up the rest of the proof of our main result as follows. First, we shall state a set of assumptions about the set S of ordered subsets of X. Then we shall show how the transference results we are aiming for follow from these assumptions. Then over the next few sections we shall show how to prove these assumptions for a large class of sets S.

The reason for doing things this way is twofold. First, it splits the proof into a deterministic part (the part we do now) and a probabilistic part (verifying the assumptions). Secondly, it splits the proof into a part that is completely general (again, the part we do now) and a part that depends more on the specific set S. Having said that, when it comes to verifying the assumptions, we do not

do so for *individual* sets S. Rather, we identify two broad classes of set S that between them cover all the problems that have traditionally interested people. This second shift, from the general to the particular, will not be necessary until Section 7. For now, the argument remains quite general.

Suppose now that $\mu_1, \ldots, \mu_m$ are measures on a finite set X and $\mu = m^{-1}(\mu_1 + \cdots + \mu_m)$. In subsequent sections, we will take $\mu_1, \ldots, \mu_m$ to be the associated measures of random sets $U_1, \ldots, U_m$, each distributed as X_p, but for now we will continue to work deterministically. We shall be particularly interested in the following four properties that such a sequence of measures may have.

Four key properties.

P0. $\|\mu_i\|_1 = 1 + o(1)$ *for each i, where $o(1) \to 0$ as $|X| \to \infty$.*

P1. $\| *_1 (\mu_{i_2}, \ldots, \mu_{i_k}) - \circ_1(\mu_{i_2}, \ldots, \mu_{i_k})\|_1 \leq \eta$ *whenever $i_2, \ldots, i_k$ are distinct integers between 1 and*
   m.

P2. $\|*_j(1, 1, \ldots, 1, \mu_{i_{j+1}}, \ldots, \mu_{i_k})\|_\infty \leq 2$ *whenever $j \geq 2$ and $i_{j+1}, \ldots, i_k$ are distinct integers between*
   1 and m.

P3. $|\langle \mu - 1, \xi \rangle| < \lambda$ *whenever ξ is a product of at most d basic anti-uniform functions from $\Phi_{\mu,1}$.*

In the remainder of this section, we will prove that if $\mu_1, \ldots, \mu_m$ satisfy these four properties, then any robust density theorem or colouring theorem also holds relative to the measure μ. To prove this for density statements, we first need a simple lemma showing that any density theorem implies an equivalent functional formulation. For convenience, we will assume that each set in S consists of distinct elements from X.

Lemma 4.4. *Let k be an integer and $\rho, \beta, \epsilon > 0$ be real numbers. Let X be a sufficiently large finite set and let S be a collection of ordered subsets of X, each of size k and with no repeated elements. Suppose that for every subset B of X of size at least $\rho|X|$ there are at least $\beta|S|$ elements $(s_1, \ldots, s_k)$ of S such that $s_i \in B$ for each i. Let g be a function on X such that $0 \leq g \leq 1$ and $\|g\|_1 \geq \rho + \epsilon$. Then*

$$\mathbb{E}_{s \in S} g(s_1) \ldots g(s_k) \geq \beta - \epsilon.$$

Proof. Let us choose a subset B of X randomly by choosing each $x \in X$ with probability $g(x)$, with the choices independent. The expected number of elements of B is $\sum_x g(x) \geq (\rho + \epsilon)|X|$ and therefore, by applying standard large deviation inequalities, one may show that if $|X|$ is sufficiently large the probability that $|B| < \rho|X|$ is at most ϵ. Therefore, with probability at least $1 - \epsilon$ there are at least $\beta|S|$ elements s of S such that $s_i \in B$ for every i. It follows that the expected number of such sequences is at least $\beta|S|(1 - \epsilon) \geq (\beta - \epsilon)|S|$. But each sequence s has probability $g(s_1) \ldots g(s_k)$ of belonging to B, so the expected number is also $\sum_{s \in S} g(s_1) \ldots g(s_k)$, which proves the lemma. $\square$

Note that the converse to the above result is trivial (and does not need an extra ϵ), since if B is a set of density ρ, then the characteristic function of B has L_1-norm ρ.

We remark here that the condition that no sequence in S should have repeated elements is not a serious restriction. For one thing, all it typically does is rule out degenerate cases (such as arithmetic progressions with common difference zero) that do not interest us. Secondly, these degenerate cases tend to be sufficiently infrequent that including them would have only a very small effect on the constants. The reason we do not allow them is that it makes the proof neater.

19

With Lemma 4.4 in hand, we are now ready to prove that a transference principle holds for density theorems.

Theorem 4.5. *Let k be a positive integer and let $\rho, \beta, \epsilon > 0$ be real numbers. Let X be a finite set and let S be a homogeneous collection of ordered subsets of X, each of size k and having no repeated elements. Suppose that for every subset B of X of size at least $\rho|X|$ there are at least $\beta|S|$ elements $(s_1, \ldots, s_k)$ of S such that $s_i \in B$ for each i. Then there are positive constants η and λ and positive integers d and m with the following property. If $\mu_1, \ldots, \mu_m$ are such that P0, P1, P2 and P3 hold for the constants η, λ and d, $\mu = m^{-1}(\mu_1 + \cdots + \mu_m)$, and $|X|$ is sufficiently large, then $\mathbb{E}_{s \in S} f(s_1) \ldots f(s_k) \geq \beta - \epsilon$ for every function f such that $0 \leq f \leq \mu$ and $\mathbb{E}_x f(x) \geq \rho + \epsilon$.*

Proof. To begin, we apply Lemma 4.4 with $\frac{\epsilon}{2}$ to conclude that if $|X|$ is sufficiently large and g is any function on X with $0 \leq g \leq 1$ and $\|g\|_1 \geq \rho + \frac{\epsilon}{2}$, then

$$\mathbb{E}_{s \in S} g(s_1) \ldots g(s_k) \geq \beta - \frac{\epsilon}{2}.$$

For each function h, let $\|h\|$ be defined to be the maximum of $|\langle h, \phi \rangle|$ over all basic anti-uniform functions $\phi \in \Phi_{\mu,1}$. Let $\eta = \frac{\epsilon}{10}$. We claim that, given f with $0 \leq f \leq \mu$, there exists a g with $0 \leq g \leq 1$ such that $\|(1 + \frac{\epsilon}{4})^{-1} f - g\| \leq \eta/k$. Equivalently, this shows that $|\langle (1 + \frac{\epsilon}{4})^{-1} f - g, \phi \rangle| \leq \eta/k$ for every $\phi \in \Phi_{\mu,1}$. We will prove this claim in a moment. However, let us first note that it is a sufficient condition to imply that

$$\mathbb{E}_{s \in S} f(s_1) \ldots f(s_k) \geq \beta - \epsilon$$

whenever $0 \leq f \leq \mu$ and $\mathbb{E}_x f(x) \geq \rho + \epsilon$. Let $m = 2k^3/\eta$ and write $(1 + \frac{\epsilon}{4})^{-1} f$ as $m^{-1}(f_1 + \cdots + f_m)$ with $0 \leq f_i \leq \mu_i$. Corollary 3.7, together with P1 and P2, then implies that $\mathbb{E}_x g(x) \geq (1 + \frac{\epsilon}{4})^{-1} \mathbb{E}_x f(x) - \eta/k$ and that

$$\left| \mathbb{E}_{i_1, \ldots, i_k \in \{1, \ldots, m\}} \langle f_{i_1}, \circ_1(f_{i_2}, \ldots, f_{i_k}) \rangle - \langle g, *_1(g, g, \ldots, g) \rangle \right| \leq 4\eta.$$

Since $\eta/k < \epsilon/8$, $(1 + \frac{\epsilon}{4})^{-1} \geq 1 - \frac{\epsilon}{4}$ and $1 + o(1) \geq \mathbb{E}_x f(x) \geq \rho + \epsilon$,

$$\mathbb{E}_x g(x) \geq \left(1 + \frac{\epsilon}{4}\right)^{-1} \mathbb{E}_x f(x) - \eta/k \geq \rho + \epsilon - \frac{\epsilon}{4} - \frac{\epsilon}{8} - o(1) \geq \rho + \frac{\epsilon}{2},$$

for $|X|$ sufficiently large, so our assumption about g implies that $\langle g, *_1(g, g, \ldots, g) \rangle \geq \beta - \frac{\epsilon}{2}$. Since in addition $8\eta < \epsilon$, we can deduce the inequality $\mathbb{E}_{i_1, \ldots, i_k} \langle f_{i_1}, \circ_1(f_{i_2}, \ldots, f_{i_k}) \rangle \geq \beta - \epsilon$, which, since the capped convolution is smaller than the standard convolution, implies that

$$\mathbb{E}_{s \in S} f(s_1) \ldots f(s_k) = \langle f, *_1(f, f, \ldots, f) \rangle \geq \mathbb{E}_{i_1, \ldots, i_k} \langle f_{i_1}, \circ_1(f_{i_2}, \ldots, f_{i_k}) \rangle \geq \beta - \epsilon.$$

It remains to prove that for any f with $0 \leq f \leq \mu$, there exists a g with $0 \leq g \leq 1$ such that $\|(1 + \frac{\epsilon}{4})^{-1} f - g\| \leq \eta/k$. An application of Lemma 2.5 tells us that if $\langle \mu - 1, \psi_+ \rangle < \frac{\epsilon}{4}$ for every function ψ with $\|\psi\|^* \leq k\eta^{-1}$, then this will indeed be the case. Now let us try to find a sufficient condition for this. First, if $\|\psi\|^* \leq k\eta^{-1}$, then Lemma 4.1 implies that ψ is contained in $k\eta^{-1}$ times the convex hull of $\Phi \cup \{-\Phi\}$, where Φ is the set of all basic anti-uniform functions. Since functions in $\Phi \cup \{-\Phi\}$ take values in $[-2, 2]$, we can apply Lemma 4.2 to find constants d and M and a function ω that can be written as M times a convex combination of products of at most d functions from $\Phi \cup \{-\Phi\}$ such that $\|\psi_+ - \omega\|_\infty \leq \epsilon/20$. Hence, for such an ω,

$$\langle \mu - 1, \psi_+ - \omega \rangle \leq \|\mu - 1\|_1 \|\psi_+ - \omega\|_\infty \leq (2 + o(1))\frac{\epsilon}{20} < \frac{\epsilon}{8},$$

for $|X|$ sufficiently large. From this it follows that if $|\langle \mu - 1, \xi \rangle| < \epsilon/8M$ whenever ξ is a product of at most d functions from $\Phi \cup \{-\Phi\}$, then

$$\langle \mu - 1, \psi_+ \rangle = \langle \mu - 1, \omega \rangle + \langle \mu - 1, \psi_+ - \omega \rangle < \epsilon/8 + \epsilon/8 = \epsilon/4.$$

Therefore, applying P3 with d and $\lambda = \epsilon/8M$ completes the proof. $\qquad \square$

To prove a corresponding theorem for colouring problems, we will again need a lemma saying that colouring theorems always have a functional reformulation.

Lemma 4.6. *Let k, r be positive integers and let $\beta > 0$ be a real number. Let X be a finite set and let S be a collection of ordered subsets of X, each of size k and having no repeated elements. Suppose that for every r-colouring of X there are at least $\beta|S|$ elements $(s_1, \ldots, s_k)$ of S such that each s_i has the same colour. Let $g_1, \ldots, g_r$ be functions from X to $[0, 1]$ such that $g_1 + \cdots + g_r = 1$. Then*

$$\mathbb{E}_{s \in S} \sum_{i=1}^{r} g_i(s_1) \ldots g_i(s_k) \geq \beta.$$

Proof. Define a random r-colouring of X as follows. For each $x \in X$, let x have colour i with probability $g_i(x)$, and let the colours be chosen independently. By hypothesis, the number of monochromatic sequences is at least $\beta|S|$, regardless of what the colouring is. But the expected number of monochromatic sequences is $\sum_{s \in S} \sum_{i=1}^{r} g_i(s_1) \ldots g_i(s_k)$, so the lemma is proved. $\qquad \square$

We actually need a slightly stronger conclusion than the one we have just obtained. However, if S is homogeneous then it is an easy matter to strengthen the above result to what we need.

Lemma 4.7. *Let k, r be positive integers and let $\beta > 0$ be a real number. Let X be a finite set and let S be a homogeneous collection of ordered subsets of X, each of size k and having no repeated elements. Suppose that for every r-colouring of X there are at least $\beta|S|$ elements $(s_1, \ldots, s_k)$ of S such that each s_i has the same colour. Then there exists $\delta > 0$ with the following property. If $g_1, \ldots, g_r$ are any r functions from X to $[0, 1]$ such that $g_1(x) + \cdots + g_r(x) \geq 1/2$ for at least $(1 - \delta)|X|$ values of x, then*

$$\mathbb{E}_{s \in S} \sum_{i=1}^{r} g_i(s_1) \ldots g_i(s_k) \geq 2^{-(k+1)} \beta.$$

Proof. Let Y be the set of x such that $g_1(x) + \cdots + g_r(x) < 1/2$. Then we can find functions $h_1, \ldots, h_r$ from X to $[0, 1]$ such that $h_1 + \cdots + h_r = 1$ and $h_i(x) \leq 2g_i(x)$ for every $x \in X \setminus Y$. By the previous lemma, we know that

$$\mathbb{E}_{s \in S} \sum_{i=1}^{r} h_i(s_1) \ldots h_i(s_k) \geq \beta.$$

Let T be the set of sequences $s \in S$ such that $s_i \in Y$ for at least one i. Since S is homogeneous, for each i the set of s such that $s_i \in Y$ has size $|S||Y|/|X| \leq \delta|S|$. Therefore, $|T| \leq k\delta|S|$. It follows that

$$
\begin{aligned}
\sum_{s \in S} \sum_{i=1}^{r} g_i(s_1) \ldots g_i(s_k) \quad &\geq \quad \sum_{s \in S \setminus T} \sum_{i=1}^{r} g_i(s_1) \ldots g_i(s_k) \\
&\geq \quad 2^{-k} \sum_{s \in S} \sum_{i=1}^{r} h_i(s_1) \ldots h_i(s_k) - |T| \\
&\geq \quad (2^{-k}\beta - k\delta)|S|.
\end{aligned}
$$

21

Thus, the lemma is proved if we take $\delta = 2^{-(k+1)}\beta/k$. $\qquad\square$

We now prove our main transference principle for colouring theorems. The proof is similar to that of Theorem 4.5 and reduces to the same conditions, but we include the proof for completeness.

Theorem 4.8. *Let k, r be positive integers and $\beta > 0$ be a real number. Let X be a finite set and let S be a homogeneous collection of ordered subsets of X, each of size k and with no repeated elements. Suppose that for every r-colouring of X there are at least $\beta|S|$ elements $(s_1, \dots, s_k)$ of S such that each s_i has the same colour. Then there are positive constants η and λ and positive integers d and m with the following property. If $\mu_1, \dots, \mu_m$ are such that P0, P1, P2 and P3 hold for the constants η, λ and d, $\mu = m^{-1}(\mu_1 + \dots + \mu_m)$, and $|X|$ is sufficiently large, then $\mathbb{E}_{s \in S} \sum_{i=1}^{r} f_i(s_1) \dots f_i(s_k) \geq 2^{-(k+2)}\beta$ for every sequence of functions $f_1, \dots, f_r$ such that $0 \leq f_i \leq \mu$ for each i and $\sum_{i=1}^{r} f_i = \mu$.*

Proof. An application of Lemmas 4.6 and 4.7 tells us that there exists $\delta > 0$ with the following property. If $g_1, \dots, g_r$ are any r functions from X to $[0, 1]$ such that $g_1(x) + \dots + g_r(x) \geq 1/2$ for at least $(1 - \delta)|X|$ values of x, then

$$\mathbb{E}_{s \in S} \sum_{i=1}^{r} g_i(s_1) \dots g_i(s_k) \geq 2^{-(k+1)}\beta.$$

Again we define the norm $\|.\|$ by taking $\|h\|$ to be the maximum of $|\langle h, \phi \rangle|$ over all basic anti-uniform functions $\phi \in \Phi_{\mu,1}$. Let η be such that $8\eta r < \min(\delta, 2^{-(k+1)}\beta)$. We claim that, given functions $f_1, \dots, f_r$ with $0 \leq f_i \leq \mu$ and $\sum_{i=1}^{r} f_i = \mu$, there are functions g_i such that $0 \leq g_i \leq 1$, $g_1 + \dots + g_r \leq 1$ and $\|(1 + \frac{\delta}{4})^{-1} f_i - g_i\| \leq \eta/k$. Equivalently, this means that $|\langle (1 + \frac{\delta}{4})^{-1} f_i - g_i, \phi \rangle| \leq \eta/k$ for every i and every $\phi \in \Phi_{\mu,1}$. We will return to the proof of this statement. For now, let us show that it implies

$$\mathbb{E}_{s \in S} \sum_{i=1}^{r} f_i(s_1) \dots f_i(s_k) \geq 2^{-(k+2)}\beta.$$

Let $m = 2k^3/\eta$ and write $(1 + \frac{\delta}{4})^{-1} f_i$ as $m^{-1}(f_{i,1} + \dots + f_{i,m})$ with $0 \leq f_{i,j} \leq \mu_j$. Corollary 3.7, together with P1 and P2, then implies that $\mathbb{E}_x g_i(x) \geq (1 + \frac{\delta}{4})^{-1} \mathbb{E}_x f_i(x) - \eta/k$ and that

$$\left| \mathbb{E}_{j_1, \dots, j_k \in \{1, \dots, m\}} \langle f_{i,j_1}, \circ_1(f_{i,j_2}, \dots, f_{i,j_k}) \rangle - \langle g_i, *_1(g_i, g_i, \dots, g_i) \rangle \right| \leq 4\eta.$$

Suppose that there were at least $\delta|X|$ values of x for which $\sum_{i=1}^{r} g_i(x) < \frac{1}{2}$. Then this would imply that

$$\mathbb{E}_{x \in X} \sum_{i=1}^{r} g_i(x) < \frac{1}{2}\delta + (1 - \delta) \leq 1 - \frac{\delta}{2}.$$

But $\mathbb{E}_x g_i(x) \geq (1 + \frac{\delta}{4})^{-1} \mathbb{E}_x f_i(x) - \eta/k$. Therefore, adding over all i, we have, since $\eta \leq \delta/8r$ and $(1 + \frac{\delta}{4})^{-1} \geq 1 - \frac{\delta}{4}$, that

$$\sum_{i=1}^{r} \mathbb{E}_{x \in X} g_i(x) \geq \left(1 + \frac{\delta}{4} \right)^{-1} (1 + o(1)) - \frac{\eta r}{k} \geq 1 - \frac{\delta}{2},$$

for $|X|$ sufficiently large, a contradiction. Our assumption about the g_i therefore implies the inequality $\sum_{i=1}^{r} \langle g_i, *_1(g_i, g_i, \dots, g_i) \rangle \geq 2^{-(k+1)}\beta$. Since $8r\eta < 2^{-(k+1)}\beta$, we can deduce the inequality

$$\sum_{i=1}^{r} \mathbb{E}_{j_1, \dots, j_k} \langle f_{i,j_1}, \circ_1(f_{i,j_2}, \dots, f_{i,j_k}) \rangle \geq 2^{-(k+2)}\beta,$$

22

which, since the capped convolution is smaller than the standard convolution, implies that

$$\mathbb{E}_{s \in S} \sum_{i=1}^{r} f_i(s_1) \dots f_i(s_k) = \sum_{i=1}^{r} \langle f_i, *_1(f_i, f_i, \dots, f_i) \rangle \geq \sum_{i=1}^{r} \mathbb{E}_{j_1, \dots, j_k} \langle f_{i,j_1}, \circ_1(f_{i,j_2}, \dots, f_{i,j_k}) \rangle \geq 2^{-(k+2)} \beta.$$

As in Theorem 4.5, we have proved our result conditional upon an assumption, this time that for any functions $f_1, \dots, f_r$ with $0 \leq f_i \leq \mu$ and $\sum_{i=1}^{r} f_i = \mu$, there are functions g_i such that $0 \leq g_i \leq 1$, $g_1 + \dots + g_r \leq 1$ and $\|(1 + \frac{\delta}{4})^{-1} f_i - g_i\| \leq \eta/k$. An application of Lemma 2.6 tells us that if $\langle \mu - 1, (\max_{1 \leq i \leq r} \psi_i)_+ \rangle < \delta/4$ for every collection of functions ψ_i with $\|\psi_i\|^* \leq k\eta^{-1}$, then this will indeed be the case. By Lemma 4.1, each ψ_i is contained in $k\eta^{-1}$ times the convex hull of $\Phi \cup \{-\Phi\}$, where Φ is the set of all basic anti-uniform functions. Since functions in $\Phi \cup \{-\Phi\}$ take values in $[-2, 2]$, we can apply Lemma 4.3 to find constants d and M and a function ω that can be written as M times a convex combination of products of at most d functions from $\Phi \cup \{-\Phi\}$, such that $\|(\max_{1 \leq i \leq r} \psi_i)_+ - \omega\|_\infty \leq \delta/20$. From this it follows that if $|X|$ is sufficiently large and $|\langle \mu - 1, \xi \rangle| < \delta/8M$ whenever ξ is a product of at most d functions from $\Phi \cup \{-\Phi\}$, then $\langle \mu - 1, (\max_{1 \leq i \leq r} \phi_i)_+ \rangle < \delta/4$. Therefore, applying P3 with d and $\lambda = \delta/8M$ proves the theorem. $\qquad \square$

Finally, we would like to talk a little about structure theorems. To motivate the result that we are about to state, let us begin by giving a very brief sketch of how to prove a sparse version of the triangle removal lemma. (For a precise statement, see Conjecture 1.7 in the introduction, and the discussion preceding it.)

The dense version of the lemma states that if a dense graph has almost no triangles, then it is possible to remove a small number of edges in order to make it triangle free. To prove this, one first applies Szemerédi's regularity lemma [60] to the graph, and then removes all edges from pairs that are sparse or irregular. Because sparse pairs contain few edges, and very few pairs are irregular, not many edges are removed. If a triangle is left in the resulting graph, then each edge of the triangle belongs to a dense regular pair, and then a simple lemma can be used to show that there must be many triangles in the graph. Since we are assuming that there are very few triangles in the graph, this is a contradiction.

The sparse version of the lemma states that essentially the same result holds in a sparse random graph, given natural interpretations of phrases such as "almost no triangles". If a random graph with n vertices has edge probability p, then the expected number of (labelled) triangles is approximately $p^3 n^3$, and the expected number of (labelled) edges is pn^2. Therefore, the obvious statement to try to prove, given a random graph G_0 with edge probability p, is this: for every $\delta > 0$ there exists $\epsilon > 0$ such that if G is any subgraph of G_0 that contains at most $\epsilon p^3 n^3$ triangles, then it is possible to remove at most δpn^2 edges from G and end up with no triangles.

How might one prove such a statement? The obvious idea is to use the transference methods explained earlier to find a $[0, 1]$-valued function g defined on pairs of vertices (which we can think of as a weighted graph) that has similar triangle-containing behaviour to G. For the sake of discussion, let us suppose that g is in fact the characteristic function of a graph and let us call that graph Γ (later, in Corollary 9.7, we will show that such a reduction is always possible).

If Γ has similar behaviour to G, then Γ contains very few triangles, which is promising. So we apply the dense triangle removal lemma in order to get rid of all triangles. But what does that tell us about G? The edges we removed from Γ did not belong to G. And in any case, how do we use an

approximate statement (that G and Γ have similar triangle-containing behaviour) to obtain an *exact* conclusion (that G with a few edges removed has *no* triangles at all)?

The answer is that we removed edges from Γ in "clumps". That is, we took pairs (U, V) of vertex sets (given by cells of the Szemerédi partition) and removed all edges linking U to V. So the natural way of removing edges from G is to remove the same clumps that we removed from Γ. After that, the idea is that if G contains a triangle then it belongs to clumps that were not removed, which means that Γ must contain a triple of dense regular clumps, and therefore many triangles, which implies that G must also contain many triangles, a contradiction.

For this to work, it is vital that if a clump contains a very small proportion of the edges of Γ, then it should also contain a very small proportion of the edges of G. More generally, the density of G in a set of the form $U \times V$ should be about p times the density of Γ in the same set. Thus, we need a result that allows us to approximate a function by one with a similar triangle count, but we also need the new function to have similar densities inside every set of the form $U \times V$ when U and V are reasonably large.

In the case of hypergraphs, we need a similar but more complicated statement. The precise nature of the complexity is, rather surprisingly, not too important: the main point is that we shall need to approximate a function dominated by a sparse random measure by a bounded function that has a similar simplex count and similar densities inside all the sets from some set system that is not too large.

In order to state the result precisely, we make the following definition.

Definition 4.9. *Suppose that we have a finite set X and suppose that $\Phi_{\mu,1}$ is a collection of basic anti-uniform functions derived from a collection S of ordered subsets of X and a sequence of measures $\mu = (\mu_1, \ldots, \mu_m)$. Then, given a collection of subsets $\mathcal{V}$ of X, we define the set of basic anti-uniform functions $\Phi_{\mu,1}(\mathcal{V})$ to be $\Phi_{\mu,1} \cup \{\chi_V : V \in \mathcal{V}\}$, where χ_V is the characteristic function of the set V.*

We also need to modify the third of the key properties, so as to take account of the set system $\mathcal{V}$.

P3′. $|\langle \mu - 1, \xi \rangle| < \lambda$ whenever ξ is a product of at most d basic anti-uniform functions from $\Phi_{\mu,1}(\mathcal{V})$.

Our main abstract result regarding the transfer of structural theorems is the following. It says that not only do the functions f and g reflect one another in the sense that they have similar subset counts, but they may be chosen to have similar densities inside all the sets V from a collection $\mathcal{V}$. The proof, which we omit, is essentially the same as that of Theorem 4.5: the only difference is that the norm is now defined in terms of $\Phi_{\mu,1}(\mathcal{V})$, which gives us the extra information that $|\langle f, \chi_V \rangle - \langle g, \chi_V \rangle| \leq \|f - g\|$ for every $V \in \mathcal{V}$ and hence the extra conclusion at the end.

Theorem 4.10. *Let k be a positive integer and $\epsilon > 0$ a constant. Let X be a finite set, S a homogeneous collection of ordered subsets of X, each of size k, and $\mathcal{V}$ a collection of subsets of X. Then there are positive constants η and λ and positive integers d and m with the following property. If $\mu_1, \ldots, \mu_m$ are such that P0, P1, P2 and P3′ hold for the constants η, λ and d, then, for $|X|$ sufficiently large, the following holds for $\mu = m^{-1}(\mu_1 + \cdots + \mu_m)$: whenever $0 \leq f \leq \mu$, there exists g with $0 \leq g \leq 1$ such that*

$$\mathbb{E}_{s \in S} f(s_1) \ldots f(s_k) \geq \mathbb{E}_{s \in S} g(s_1) \ldots g(s_k) - \epsilon$$

and, for all $V \in \mathcal{V}$,

$$|\mathbb{E}_{x \in V} f(x) - \mathbb{E}_{x \in V} g(x)| \leq \epsilon \frac{|X|}{|V|}.$$

We remark that the second part of the conclusion can be rewritten as

$$|\mathbb{E}_x \chi_V(x) f(x) - \mathbb{E}_x \chi_V(x) g(x)| \leq \epsilon,$$

which is precisely the statement that $|\langle f, \chi_V \rangle - \langle g, \chi_V \rangle| \leq \epsilon$.

Note that for P3′ to have a chance of holding, we cannot have too many sets in the collection $\mathcal{V}$. However, we can easily have enough sets for our purposes. For instance, the collection of pairs of vertex sets in a graph with n vertices has size 4^n; this is far smaller than the number of graphs on n vertices, which is exponential in n^2. More generally, an important role in the hypergraph regularity lemma is played by k-uniform hypergraphs H formed as follows: take a $(k-1)$-uniform hypergraph K and for each set E of size k, put E into H if and only if all its subsets of size $k-1$ belong to K. Since there are far fewer $(k-1)$-uniform hypergraphs than there are k-uniform hypergraphs, we have no trouble applying our result.

Since our ultimate goal is to prove a probabilistic theorem, the task that remains to us is to prove that certain random sets satisfy P0, P1, P2 and P3 (or P3′) with high probability. That this is so for P0 follows easily from Chernoff's inequality. It remains to consider P1, P2 and P3.

5 Small correlation with a fixed function

One of our main aims in this paper is to show that, with high probability, $|\langle \mu - 1, \xi \rangle| < \lambda$ for every product ξ of at most d basic anti-uniform functions, when μ is chosen randomly with suitable density. This is a somewhat complicated statement, since the set of basic anti-uniform functions depends on our random variable μ. In this section we prove a much easier result, which will nevertheless be useful to us later on: we shall show that, for any *fixed* bounded function ξ, $|\langle \mu - 1, \xi \rangle| < \lambda$ with high probability.

To prove this, we shall need a standard probabilistic estimate, Bernstein's inequality, which allows one to bound the sum of independent and not necessarily identically distributed random variables.

Lemma 5.1. *Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables. Suppose that each Y_i lies in the interval $[0, M]$. Let $S = Y_1 + Y_2 + \cdots + Y_n$. Then*

$$\mathbb{P}(|S - \mathbb{E}(S)| \geq t) \leq 2 \exp \left\{ \frac{-t^2}{2 \left(\sum \mathbb{V}(Y_j) + \frac{Mt}{3} \right)} \right\}.$$

We are now ready to prove that $\langle \mu - 1, \xi \rangle$ is bounded with high probability for any fixed bounded function ξ.

Lemma 5.2. *Let X be a finite set and let $U = X_p$. Let μ be the associated measure of U. Then, for any constants C and λ with $C \geq \lambda$ and any positive function ξ with $\|\xi\|_\infty \leq C$,*

$$\mathbb{P}(|\langle \mu - 1, \xi \rangle| \geq \lambda) \leq 2e^{-\lambda^2 p |X|/3C^2}.$$

Proof. For each $x \in X$, $\mu(x)\xi(x)$ is a random variable that takes values in the interval $[0, p^{-1}C]$. The expectation of $\mu(x)\xi(x)$ is $\xi(x)$, so the expectation of $\langle \mu - 1, \xi \rangle$ is 0. Also, the variance of $\mu(x)\xi(x)$ is at most $\mathbb{E}(\mu(x)^2\xi(x)^2)$, which is $\xi(x)^2 p^{-1}$, which is at most $C^2 p^{-1}$.

Let $S = \sum_x \mu(x)\xi(x)$. Then the probability that $|\langle \mu - 1, \xi \rangle| \geq \lambda$ equals the probability that $|S - \mathbb{E}S| \geq \lambda|X|$. Therefore, by Bernstein's inequality,

$$
\begin{aligned}
\mathbb{P}(|\langle \mu - 1, \xi \rangle| \geq \lambda) &\leq 2 \exp \left\{ \frac{-(\lambda|X|)^2}{2\left(C^2 p^{-1}|X| + C\lambda p^{-1}|X|/3\right)} \right\} \\
&= 2 \exp \left\{ \frac{-\lambda^2 p|X|}{2\left(C^2 + C\lambda/3\right)} \right\} \\
&\leq 2 \exp \left\{ -\lambda^2 p|X|/3C^2 \right\},
\end{aligned}
$$

where to prove the second inequality we used the assumption that $C \geq \lambda$. $\qquad \square$

Before we move on to the next section, it will be helpful to state Chernoff's inequality, the standard estimate for the tails of the binomial distribution. As we have already noted, P0 is a straightforward consequence of this lemma.

Lemma 5.3. *Let $0 < p \leq 1$ and $0 < \delta \leq \frac{1}{2}$ be real numbers and X a finite set. Then*

$$
\mathbb{P}(||X_p| - p|X|| \geq \delta p|X|) \leq 2e^{-\delta^2 p|X|/4}.
$$

6 The set of basic anti-uniform functions has few extreme points

A slightly inaccurate description of what we are going to do next is that we shall show that if P1 and P2 hold with high probability, then so does P3. In order to understand how and why what we shall actually do differs from this, it is important to understand the difficulty that we now have to overcome. The result of the previous section tells us that for a random measure μ and any given function ξ, $|\langle \mu - 1, \xi \rangle|$ is bounded with high probability. We now need to show that this is the case for *all* functions ξ that are products of at most d basic anti-uniform functions. As we have already commented, this is tricky, because which functions are basic anti-uniform functions depends on μ.

To get a clearer notion of the problem, let us look at a subcase of the general fact that we are trying to prove, by thinking how we might try to show that, with high probability, $\langle \mu_1 - 1, \circ_1(f_2, \ldots, f_k) \rangle$ is small whenever $0 \leq f_i \leq \mu_i$ for $i = 2, 3, \ldots, k$. That is, for the time being we shall concentrate on basic anti-uniform functions themselves rather than on products of such functions.

A question that will obviously be important to us is the following: for how many choices of functions $f_2, \ldots, f_k$ do we need to establish that $|\langle \mu_1 - 1, \circ_1(f_2, \ldots, f_k) \rangle|$ is small? At first glance, the answer might seem to be infinitely many, but one quickly realizes that a small uniform perturbation to the functions $f_2, \ldots, f_k$ does not make much difference to $\langle \mu_1 - 1, \circ_1(f_2, \ldots, f_k) \rangle$. So it will be enough to look at some kind of net of the functions.

However, even this observation is not good enough, since the number of functions in a net will definitely be at least exponentially large in $p|X|$. Although the probability we calculated in the previous section is exponentially small in $p|X|$, the constant is small, whereas the constant involved in the size of a net will not be small. So it looks as though there are too many events to consider.

It is clear that the only way round this problem is to prove that the set of basic anti-uniform functions $\circ_1(f_2, \ldots, f_k)$ is somehow smaller than expected. And once one thinks about this for a bit, one realizes that this may well be the case. So far, we have noted that $\circ_1(f_2, \ldots, f_k)$ is not much affected by small uniform perturbations to the functions f_i. However, an important theme in additive combinatorics is that convolutions tend to be robust under a much larger class of perturbations: roughly speaking, a "quasirandom" perturbation of one of the f_i is likely to have little effect on $\circ_1(f_2, \ldots, f_k)$.

It is not immediately obvious how to turn this vague idea into a precise one, so for a moment let us think more abstractly. We have a class Γ of functions, and a function ν, and we would like to prove that $\langle \nu, \phi \rangle$ is small for every $\phi \in \Gamma$. To do this, we would like to identify a much smaller class of functions Δ such that if $\langle \nu, \psi \rangle$ is small for every $\psi \in \Delta$ then $\langle \nu, \phi \rangle$ is small for every $\phi \in \Gamma$. The following very simple lemma tells us a sufficient (and also in fact necessary) condition on Δ for us to be able to make this deduction.

Lemma 6.1. *Let Γ and Δ be two closed sets of functions from X to $\mathbb{R}$ and suppose that both are centrally symmetric. Then the following two statements are equivalent:*

(i) For every function ν, $\max\{|\langle \nu, \phi \rangle| : \phi \in \Gamma\} \leq \max\{|\langle \nu, \psi \rangle| : \psi \in \Delta\}$.

(ii) Γ is contained in the convex hull of Δ.

Proof. The statement we shall use is just the easy direction of this equivalence, which is that (ii) implies (i). To see this, let $\phi \in \Gamma$. Then we can write ϕ as a convex combination $\sum_i \lambda_i \psi_i$ of elements of Δ, and that implies that $|\langle \nu, \phi \rangle| \leq \sum_i \lambda_i |\langle \nu, \psi_i \rangle|$. If $|\langle \nu, \psi \rangle| \leq t$ for every $\psi \in \Delta$, then this is at most t, which proves (i), since ν and ϕ were arbitrary.

Now let us suppose that Γ is not contained in the convex hull of Δ, and let ϕ be an element of Γ that does not belong to this convex hull. Then the Hahn-Banach theorem and the fact that Δ is closed and centrally symmetric guarantee the existence of a function ν such that $\langle \nu, \phi \rangle > 1$, but $|\langle \nu, \psi \rangle| \leq 1$ for every $\psi \in \Delta$, which contradicts (i). $\square$

The reason Lemma 6.1 is useful is that it gives us a strategy for proving that $|\langle \mu - 1, \xi \rangle|$ is small for all products of at most d basic anti-uniform functions: try to show that these functions belong to the convex hull of a much smaller set. In fact, this is not quite what we shall do. Rather, we shall show that every ξ can be *approximated* by an element of the convex hull of a much smaller set. To prepare for the more elaborate statement we shall use, we need another easy lemma.

The statement of the lemma is not quite what one might expect. The reason for this is that the simplest notion of approximation, namely uniform approximation, is too much for us to hope to attain. Instead, we go for a kind of weighted uniform approximation, where we allow the functions to differ quite a lot, but only in a few specified places.

Lemma 6.2. *Let H be a non-negative function defined on X such that $\|H\|_1 \leq \epsilon$ and $\|H\|_\infty \leq R$. Let $U = X_p$ and let μ be the associated measure of U. Then, with probability at least $1 - 2\exp(-\epsilon^2 p|X|/3R^2)$, we have the estimate $|\langle \mu - 1, \phi \rangle - \langle \mu - 1, \psi \rangle| \leq 3\epsilon$ for every pair of functions ϕ and ψ such that $|\phi - \psi| \leq H$.*

Proof. The fact that $\|H\|_1 \leq \epsilon$ implies that $|\langle 1, \phi \rangle - \langle 1, \psi \rangle| \leq \epsilon$ as well. Also, $|\langle \mu, \phi - \psi \rangle| \leq \langle \mu, H \rangle$. Therefore, it remains to estimate the probability that $\langle \mu, H \rangle > 2\epsilon$. Lemma 5.2 with $\lambda = \epsilon$ and $C = R$ implies that the probability that $\langle \mu - 1, H \rangle > \epsilon$ is at most $2\exp(-\epsilon^2 p|X|/3R^2)$. Therefore, with probability at least $1 - 2\exp(-\epsilon^2 p|X|/3R^2)$, $\langle \mu, H \rangle \leq 2\epsilon$. The result follows. $\square$

If we use Lemma 6.1 and Lemma 6.2 in combination, then we can show the following result.

Corollary 6.3. *Let H be a non-negative function defined on X such that $\|H\|_1 \leq \epsilon$ and $\|H\|_\infty \leq R$. Let $U = X_p$ and let μ be the associated measure of U. Let Γ and Δ be two sets of functions and suppose that for every $\phi \in \Gamma$ there exists ψ in the convex hull of Δ such that $|\phi - \psi| \leq H$. Then*

$$\max\{|\langle \mu - 1, \phi \rangle| : \phi \in \Gamma\} \leq \max\{|\langle \mu - 1, \psi' \rangle| : \psi' \in \Delta\} + 3\epsilon$$

with probability at least $1 - 2\exp(-\epsilon^2 p|X|/3R^2)$.

Proof. By Lemma 6.2, the probability is at least $1 - 2\exp(-\epsilon^2 p|X|/3R^2)$ that $|\langle \mu - 1, \phi \rangle - \langle \mu - 1, \psi \rangle| \leq 3\epsilon$ whenever $|\phi - \psi| \leq H$. By the easy direction of Lemma 6.1, $|\langle \mu - 1, \psi \rangle| \leq \max\{|\langle \mu - 1, \psi' \rangle| : \psi' \in \Delta\}$ for every ψ in the convex hull of Δ. This proves the result. $\square$

How do we define an appropriate set of functions Δ? A simple observation gets us close to the set we need, but we shall need to make a definition before we can explain it.

Definition 6.4. *Let $0 < q \leq p \leq 1$, $U = X_p$ and let $V = U_{q/p}$. Let μ be the associated measure of U and let ν be the associated measure of V considered as a set distributed as X_q. Let f be a function with $0 \leq f \leq \mu$. Then the* normalized restriction f_ν *of f to V is the function defined by taking $f_\nu(x) = (p/q)f(x)$ if $x \in V$ and 0 otherwise.*

The normalization is chosen to give us the following easy lemma. Note that the expectation below is a "probabilistic" expectation rather than a mere average over a finite set.

Lemma 6.5. *Let $U = X_p$ be a set with associated measure μ and let $V = U_{q/p}$ be a random subset of U with associated measure ν. Then, for any function $0 \leq f \leq \mu$, $f = \mathbb{E}_V f_\nu$.*

Proof. For each $x \in U$ we have

$$\mathbb{E}_V f_\nu(x) = (p/q)f(x)\mathbb{P}[x \in V] = f(x),$$

and for each $x \notin U$ we have $f(x) = \mathbb{E}_V f_\nu(x) = 0$. $\square$

This lemma expresses f as a convex combination of normalized restrictions, which is potentially useful to us, since if q/p is a small constant, then a typical contribution to this convex combination comes from a restriction to a set that is quite a lot smaller than U. That will allow us to find a small net for the set of all possible restrictions, whose convex hull can then be used to approximate the set of all possible functions f with $0 \leq f \leq \mu$.

Furthermore, we can use Lemma 6.5 to write convolutions and products of convolutions as convex combinations as well, as the next lemma easily implies. The lemma itself is very easy, so we omit the proof.

Lemma 6.6. *For $1 \leq i \leq m$, let f_i be a fixed function and let g_i be a random function such that $f_i = \mathbb{E}g_i$. Let $\kappa(f_1, \ldots, f_m)$ be a multilinear form in the functions $f_1, \ldots, f_m$. Then*

$$\kappa(f_1, \ldots, f_m) = \mathbb{E}\kappa(g_1, \ldots, g_m).$$

The rough idea, and the third main idea of the paper, is to rewrite every product of convolutions as an average of products of basic anti-uniform functions built out of normalized restrictions. Since there are "fewer" of these, we will have fewer events that need to hold. We must, however, be careful when we apply this idea. For a start, we cannot afford to let q become too small. If q is too small, then, given associated measures $\nu_2, \ldots, \nu_k$ of sets distributed as X_q, we can no longer guarantee that the convolutions $*_1(\nu_2, \cdots, \nu_k)$ are sufficiently well-behaved for our purposes. Even when we choose q to be large enough, there will still be certain rogue choices of sets. However, for q sufficiently large, we can take care of these rogue sets by averaging and showing that they make a small contribution. Thus, there is a delicate balance involved: q has to be small enough to give rise to a small class of functions, but large enough for these functions to have the properties required of them.

Another problem arises from the form of the basic anti-uniform functions. Recall that our starting point is a collection of (randomly chosen) sets $U_1, \ldots, U_m$ with associated measures $\mu_1, \ldots, \mu_m$. The basic anti-uniform functions we want to approximate are of the form $\circ_j(g, \ldots, g, f_{j+1}, \ldots, f_k)$, where $0 \leq g \leq 1$ and $0 \leq f_h \leq \mu_{i_h}$ for some sequence $(i_{j+1}, \ldots, i_k)$ of integers between 1 and m. Therefore, we must approximate capped convolutions of functions some of which are bounded above by 1 and some of which are bounded above by associated measures of sparse random sets. This creates a difficulty for us. It is still true that if $V = X_q$ is a random set with associated measure ν, then $g = \mathbb{E}_V g_\nu$, but if we exploit that fact directly, then the number of sets V that we have to consider is on the order of $\binom{|X|}{q|X|}$. Since we will take q/p to be a constant, this is much larger than $\exp(cp|X|)$ for any constant c and therefore too large to use in a probabilistic estimate given by a simple union bound. To get round this problem we shall find a much smaller set $\mathcal{V}$ such that $g = \mathbb{E}_{V \in \mathcal{V}} g_\nu$.

We shall need the following piece of notation to do this. Suppose that the elements of the set X are ordered in some arbitrary way as $x_1, \cdots, x_n$, say. Then, given a subset $V = \{x_{j_1}, \ldots, x_{j_l}\}$ of X and an integer a between 0 and $n-1$, we define the set $V + a$ to be the set formed by translating the indices by a. That is, $V + a = \{x_{j_1+a}, \ldots, x_{j_l+a}\}$ where the sums are taken modulo n. (This "translation" operation has no mathematical significance: it just turns out to be a convenient way of defining a small collection of sets.)

Let us write $\nu + a$ for the characteristic measure of $V + a$. Write $g_{\nu+a}$ for the function given by $\frac{|X|}{|V|} g(x)$ if $x \in V + a$ and 0 otherwise. A proof almost identical to that of Lemma 6.5 implies that $g = \mathbb{E}_a g_{\nu+a}$ for any function g, and in particular for any function g such that $0 \leq g \leq 1$. From this and Lemma 6.5 itself it follows that if $(W_1, \ldots, W_{j-1})$ are any subsets of X, and ω_i is the characteristic measure of W_i, then

$$*_j(g, \ldots, g, f_{j+1}, \ldots, f_k) = \mathbb{E}_{a_1, \ldots, a_{j-1}} \mathbb{E}_{V_{j+1}, \ldots, V_k} *_j (g_{\omega_1+a_1}, \ldots, g_{\omega_{j-1}+a_{j-1}}, (f_{j+1})_{\nu_{j+1}}, \ldots, (f_k)_{\nu_k}),$$

where for $h > j$ the set V_h is distributed as $(U_{i_h})_{q/p}$. There is one practical caveat, in that this identity holds when $\omega_1, \ldots, \omega_{j-1}$ are characteristic measures, but it is more natural for us to deal with associated measures. However, if we assume that $W_1, \ldots, W_{j-1}$ were chosen with probability q and $|W_i| = (1 + o(1))q|X|$, then the distinction vanishes and the identity above holds (up to a $o(1)$ term) with associated measures rather than characteristic ones.

This observation is encouraging, because it represents the convolution $*_j(g, \ldots, g, f_{j+1}, \ldots, f_k)$ as an average of convolutions from a small class of functions. However, it certainly does not solve our problems completely, since we need a statement about *capped* convolutions. Of course, it would be very surprising if it did solve our problems, since so far we have not said anything about the size of

q and the sets $W_1, \ldots, W_{j-1}$. In order to transfer the trivial observation above from convolutions to capped convolutions, we shall need q to be sufficiently large and the W_i to be "sufficiently random".

6.1 Sufficient randomness

First, let us describe two properties that are closely related to the properties P1 and P2 defined earlier, and discuss how they are related. The properties will apply to a sequence of measures $\nu_1, \ldots, \nu_{j-1}, \nu_{j+1}, \ldots, \nu_k$ and parameters $\eta > 0$ and $j \leq k$.

Q1. $\| *_j (\nu_1, \ldots, \nu_{j-1}, \nu_{j+1}, \ldots, \nu_k) - \circ_j (\nu_1, \ldots, \nu_{j-1}, \nu_{j+1}, \ldots, \nu_k) \|_1 \leq \eta$.

Q2. $\| *_j (1, \ldots, 1, \nu_{j+1}, \ldots, \nu_k) \|_\infty \leq 2$.

The main difference between these new properties and the properties P1 and P2 is that we are not quantifying over a whole set of sequences. For example, P1 is the property that Q1 holds with $j = 1$ for every sequence $(\mu_{i_2}, \ldots, \mu_{i_k})$ taken from a sequence $(\mu_1, \ldots, \mu_m)$.

A less obvious difference is that, while we are ultimately interested in obtaining properties of the measures $\mu_1, \ldots, \mu_m$, we shall deduce these from probabilistic statements about typical sequences of measures $\nu_1, \ldots, \nu_k$ chosen binomially with a smaller probability. This will be illustrated by the main result of this section.

Let us define what we mean by "sufficiently random" and then show that what we need can be obtained if Q1 holds with sufficiently high probability for suitable η. The next definition highlights the property that we want to get out of the sufficient randomness: that capped convolutions should be pretty similar to actual convolutions.

The randomness property we need of our sets W_i is roughly speaking that almost all sequences of sets that appear in the averages we consider satisfy Q1 for some small η. That will allow us to prove a statement about capped convolutions, because almost all the convolutions that appear in the average in the observation above can then be approximated by their capped counterparts. Here is the formal definition.

Definition 6.7. *Let $\eta > 0$ be a real number, let $0 < q \leq p \leq 1$ and let $W_1, \ldots, W_{j-1}$ and $Z_{j+1}, \ldots, Z_k$ be subsets of X. We say that $W_1, \ldots, W_{j-1}, Z_{j+1}, \ldots, Z_k$ are sufficiently random if $|W_h| = (1 + o(1))q|X|$ for every $h < j$, $|Z_h| = (1 + o(1))p|X|$ for every $h > j$ and, if $\omega_1, \ldots, \omega_{j-1}$ are the associated measures of $W_1, \ldots, W_{j-1}$ defined with weight q^{-1}, then the following statement holds.*

- *Let a sequence $(a_1, \ldots, a_{j-1}, V_{j+1}, \ldots, V_k)$ be chosen randomly and independently such that $a_h \in \mathbb{Z}_n$ for every $h < j$, $V_h = (Z_h)_{q/p}$ for every $h > j$ and, for each $h > j$, let ν_h be the associated measure of V_h defined with weight q^{-1}. Then the probability that the $(k-1)$-tuple $(\omega_1 + a_1, \ldots, \omega_{j-1} + a_{j-1}, \nu_{j+1}, \ldots, \nu_k)$ satisfies Q1 is at least $1 - o(|X|^{-k})$.*

Strictly speaking, the definition of sufficient randomness depends on the parameters η, p and q, but these will be clear from the context.

Our next main lemma says that if Q1 holds with sufficiently high probability, then the probability that $W_1, \ldots, W_{j-1}, Z_{j+1}, \ldots, Z_k$ are sufficiently random, if we choose them independently at random in a suitable way, is also close to 1.

Lemma 6.8. *Suppose that if $\nu_1, \ldots, \nu_k$ are the associated measures of sets $V_1, \ldots, V_k$, each chosen binomially with probability $q \geq p_0$, then property Q1 holds with probability $1 - o(|X|^{-k})$. For $1 \geq p \geq q \geq p_0$, let $W_1, \ldots, W_{j-1}$ be independent random subsets of X with each $W_h = X_q$, and let $Z_{j+1}, \ldots, Z_k$ be independent random sets with each $Z_h = X_p$. Then the probability that $W_1, \ldots, W_{j-1}, Z_{j+1}, \ldots, Z_k$ are sufficiently random is $1 - o(1)$.*

Proof. Consider the following way of choosing $k - 1$ random sets. First we choose $W_1, \ldots, W_{j-1}$ and $Z_{j+1}, \ldots, Z_k$ as in the statement of the lemma. Chernoff's inequality easily implies that with probability $1 - o(|X|^{-k})$ we have $|W_h| = (1 + o(1))q|X|$ for all $h < j$ and $|Z_h| = (1 + o(1))p|X|$ for all $h > j$. Next, we choose $a_1, \ldots, a_{j-1}, V_{j+1}, \ldots, V_k$ randomly and independently such that $a_h \in \mathbb{Z}_n$ for every $h < j$ and $V_h = (Z_h)_{q/p}$ for every $h > j$. Then the sets $W_1 + a_1, \ldots, W_{j-1} + a_{j-1}, V_{j+1}, \ldots, V_k$ are independent random sets, each distributed as X_q.

Let their associated measures be $\omega_1 + a_1, \ldots, \omega_{j-1} + a_{j-1}, \nu_{j+1}, \ldots, \nu_k$. Then, since $q \geq p_0$, our assumption tells us that this sequence of measures satisfies Q1 with probability $1 - o(|X|^{-k})$. Therefore, with probability $1 - o(1)$, when we choose the W_i and the Z_i, the probability that the sequence $\omega_1 + a_1, \ldots, \omega_{j-1} + a_{j-1}, \nu_{j+1}, \ldots, \nu_k$ satisfies Q1, conditional on that choice of the W_i and Z_i, is at least $1 - o(|X|^{-k})$. Hence, $W_1, \ldots, W_{j-1}, Z_{j+1}, \ldots, Z_k$ are sufficiently random with probability $1 - o(1)$, as claimed. $\square$

In particular, if $Z_{j+1}, \ldots, Z_k$ are binomial random subsets of X, each chosen with probability p, then, with high probability, there is a choice of sets $W_1, \ldots, W_{j-1}$ such that $W_1, \ldots, W_{j-1}, Z_{j+1}, \ldots, Z_k$ are sufficiently random.

6.2 The proof for basic anti-uniform functions

Suppose that $Lq = p \geq q \geq p_0$ for some large constant L. It will be by choosing this constant L to be large enough that we will make our trick of using normalized restrictions work. Throughout this section, we will assume that $\eta > 0$ is some parameter yet to be specified, and $Z_{j+1}, \ldots, Z_k$ is a sequence of sets, with associated measures $\zeta_{j+1}, \ldots, \zeta_k$ defined with weight p^{-1}, such that

(i) If $j = 1$, then Q1 holds for the sequence of measures $(\zeta_2, \ldots, \zeta_k)$.

(ii) If $j > 1$, then Q2 holds for the sequence of measures $(\zeta_{j+1}, \ldots, \zeta_k)$.

(iii) There exist sets $W_1, \ldots, W_{j-1}$ such that $W_1, \ldots, W_{j-1}, Z_{j+1}, \ldots, Z_k$ are sufficiently random (with parameters j and η).

In the remainder of the section, we will show that if conditions (i), (ii) and (iii) hold (with parameters j and η for suitable η), then the set of basic anti-uniform functions defined using $\zeta_{j+1}, \ldots, \zeta_k$ has a small net. To be more precise, we need some definitions. In what follows, we will write $\omega_{h,a}$ for the associated measure of $W_h + a$ (or, more accurately, the translate by a of the associated measure ω_h of W_h), where again these associated measures are defined with weight q^{-1}. We will also write $\omega'_{h,a}$ for the characteristic measure of $W_h + a$.

Definition 6.9. *Let $\Phi(\zeta_{j+1}, \ldots, \zeta_k)$ be the set of functions $\circ_j(g, \ldots, g, f_{j+1}, \ldots, f_k)$, where $0 \leq g \leq 1$ and $0 \leq f_h \leq \zeta_h$ for each $h > j$. Let $\Psi(\zeta_{j+1}, \ldots, \zeta_k)$ be the set of functions $\circ_j(f_1, \ldots, f_{j-1}, f_{j+1}, \ldots, f_k)$ such that the constituent functions f_h have the following properties. If $h < j$, then $0 \leq f_h \leq \omega'_{h,a}$ for*

31

some a, and if $h > j$ then $0 \leq f_h \leq \nu_h$, where ν_h is the associated measure, defined with weight q^{-1}, of some set $V_h \in (Z_h)_{q/p}$ such that $|V_h| \leq 2q|X|$.

We shall now show that every function in $\Phi(\zeta_{j+1}, \ldots, \zeta_k)$ can be approximated by a convex combination of functions in $\Psi(\zeta_{j+1}, \ldots, \zeta_k)$. This will be very useful to us, because $\Psi(\zeta_{j+1}, \ldots, \zeta_k)$ is a much "smaller" set than $\Phi(\zeta_{j+1}, \ldots, \zeta_k)$. However, we need to be rather careful about precisely what we mean by "can be approximated by".

Lemma 6.10. *Let L and $\alpha < 1$ be positive constants. If η is sufficiently small (depending on α) and $|X|$ is sufficiently large (depending on L and α), then there is a non-negative function H such that $\|H\|_1 \leq \alpha$, $\|H\|_\infty \leq 2$, and, for every function $\circ_j(g, \ldots, g, f_{j+1}, \ldots, f_k) \in \Phi(\zeta_{j+1}, \ldots, \zeta_k)$, there exists a function σ in the convex hull of $\Psi(\zeta_{j+1}, \ldots, \zeta_k)$ such that*

$$0 \leq \circ_j(g, \ldots, g, f_{j+1}, \ldots, f_k) - \sigma \leq H.$$

Proof. Let us choose a random function $\psi \in \Psi(\zeta_{j+1}, \ldots, \zeta_k)$ as follows. Suppose that $W_1, \ldots, W_{j-1}$ are the sets given by condition (iii) and that for each h the associated measure of W_h is ω_h (as in the definition of sufficient randomness). We start by choosing a random sequence $(\omega_1', \ldots, \omega_{j-1}', \nu_{j+1}, \ldots, \nu_k)$. Here, each ω_h' is chosen uniformly at random from the $|X|$ measures $\omega_{h,a}'$ and each ν_h is the associated measure of a set V_h, where the sets V_h are independent and distributed as $(Z_h)_{q/p}$. We then let ψ be the function

$$\circ_j(g_{\omega_1'}, \ldots, g_{\omega_{j-1}'}, (f_{j+1})_{\nu_{j+1}}, \ldots, (f_k)_{\nu_k})$$

if every V_h has size at most $2q|X|$, and the zero function otherwise. Finally, we take σ to be the expectation of ψ, which is certainly a convex combination of functions in $\Psi(\zeta_{j+1}, \ldots, \zeta_k)$.

Let us begin with the first inequality. Here we shall prove the slightly stronger result that the inequality holds even if we take $\psi = \circ_j(g_{\omega_1'}, \ldots, g_{\omega_{j-1}'}, (f_{j+1})_{\nu_{j+1}}, \ldots, (f_k)_{\nu_k})$ for all choices of V_h (rather than setting it to be zero when one of the V_h is too large).

Let T be the function from $\mathbb{R}$ to $\mathbb{R}$ defined by $T(y) = \min\{y, 2\}$ and let $S(y) = y - T(y) = \max\{y - 2, 0\}$. Then

$$
\begin{aligned}
\circ_j(g, \ldots, g, f_{j+1}, \ldots, f_k) &= T(*_j(g, \ldots, g, f_{j+1}, \ldots, f_k)) \\
&= T(\mathbb{E}(*_j(g_{\omega_1'}, \ldots, g_{\omega_{j-1}'}, (f_{j+1})_{\nu_{j+1}}, \ldots, (f_k)_{\nu_k}))),
\end{aligned}
$$

by Lemmas 6.5 and 6.6 and the fact that $g = \mathbb{E}_a g_{\omega_{h,a}'}$ (the reason we use the characteristic measures $\omega_{h,a}'$ rather than the associated measures $\omega_{h,a}$ here is so that this identity is exactly true rather than merely approximately true with high probability). On the other hand,

$$\mathbb{E}(\circ_j(g_{\omega_1'}, \ldots, g_{\omega_{j-1}'}, (f_{j+1})_{\nu_{j+1}}, \ldots, (f_k)_{\nu_k})) = \mathbb{E}T(*_j(g_{\omega_1'}, \ldots, g_{\omega_{j-1}'}, (f_{j+1})_{\nu_{j+1}}, \ldots, (f_k)_{\nu_k})).$$

Since T is a concave function, the result follows from Jensen's inequality.

Now let us define H and prove the second inequality. Since capped convolutions are smaller than convolutions and, as above, $*_j(g, \ldots, g, f_{j+1}, \ldots, f_k)) = \mathbb{E}(*_j(g_{\omega_1'}, \ldots, g_{\omega_{j-1}'}, (f_{j+1})_{\nu_{j+1}}, \ldots, (f_k)_{\nu_k}))$, the left-hand side of this inequality is at most $\mathbb{E}(*_j(g_{\omega_1'}, \ldots, g_{\omega_{j-1}'}, (f_{j+1})_{\nu_{j+1}}, \ldots, (f_k)_{\nu_k}) - \psi)$, which is $\mathbb{E}\tau$, where

$$\tau = S(*_j(g_{\omega_1'}, \ldots, g_{\omega_{j-1}'}, (f_{j+1})_{\nu_{j+1}}, \ldots, (f_k)_{\nu_k}))$$

if every V_h has size at most $2q|X|$, and $*_j(g_{\omega_1'}, \ldots, g_{\omega_{j-1}'}, (f_{j+1})_{\nu_{j+1}}, \ldots, (f_k)_{\nu_k})$ otherwise.

If every V_h has size at most $2q|X|$, then $\tau \leq S(*_j(\omega_1', \ldots, \omega_{j-1}', \nu_{j+1}, \ldots, \nu_k))$, since S is an increasing function. If there is some set V_h which is too large, then we use the bound

$$*_j(g_{\omega_1'}, \ldots, g_{\omega_{j-1}'}, (f_{j+1})_{\nu_{j+1}}, \ldots, (f_k)_{\nu_k}) \leq |X|^k *_j(1, \ldots, 1) = |X|^k,$$

which follows since $g_{\omega_h'} \leq \omega_h' \leq \frac{|X|}{|W_h|} \leq |X|$ and $(f_h)_{\nu_h} \leq \nu_h \leq (p/q)\zeta_h \leq q^{-1} \leq |X|$ for every h. Since, by Chernoff's inequality, the probability that some one of the V_h has size larger than $2q|X|$ is exponentially small in $q|X|$, the contribution of these bad terms is $o(1)$ everywhere.

We also have the trivial bound

$$\circ_j(g, \ldots, g, f_{j+1}, \ldots, f_k) - \sigma \leq 2.$$

Accordingly, if we set $H = \min\{\eta + \mathbb{E}S(*_j(\omega_1', \ldots, \omega_{j-1}', \nu_{j+1}, \ldots, \nu_k)), 2\}$, then, provided $|X|$ is sufficiently large, we have a function that satisfies the second inequality and trivially satisfies the inequality $\|H\|_\infty \leq 2$.

It remains to bound $\|H\|_1$. Let $\eta = \alpha/4$. The sufficient randomness assumption tells us that when we choose our random sequence $(\omega_1, \ldots, \omega_{j-1}, \nu_{j+1}, \ldots, \nu_k)$, the probability that it satisfies Q1 is $1 - o(|X|^{-k})$. Since there are at most $|X|^k$ ways of choosing $\omega_1, \ldots, \omega_{j-1}$, it follows that with probability $1 - o(1)$, every single such choice results in a sequence that satisfies Q1. That is, we have the inequality

$$\| *_j(\omega_{1,a_1}, \ldots, \omega_{j-1,a_{j-1}}, \nu_{j+1}, \ldots, \nu_k) - \circ_j(\omega_{1,a_1}, \ldots, \omega_{j-1,a_{j-1}}, \nu_{j+1}, \ldots, \nu_k)\|_1 \leq \eta.$$

The condition that $|W_i| = (1 + o(1))q|X|$ implies that, for every $1 \leq i \leq k$ and every $a \in |X|$, $\omega_{i,a} = (1 + o(1))\omega_{i,a}'$. Therefore, for $|X|$ sufficiently large,

$$\| *_j(\omega_{1,a_1}', \ldots, \omega_{j-1,a_{j-1}}', \nu_{j+1}, \ldots, \nu_k) - \circ_j(\omega_{1,a_1}', \ldots, \omega_{j-1,a_{j-1}}', \nu_{j+1}, \ldots, \nu_k)\|_1 \leq 2\eta,$$

for any $\nu_{j+1}, \ldots, \nu_k$ such that every choice of $\omega_1, \ldots, \omega_{j-1}$ yields a sequence that satisfies Q1.

If $\nu_{j+1}, \ldots, \nu_k$ are such that there exists $(a_1, \ldots, a_{j-1})$ for which $(\omega_{1,a_1}, \ldots, \omega_{j-1,a_{j-1}}, \nu_{j+1}, \ldots, \nu_k)$ does not satisfy Q1, then we use a "trivial" bound instead. For each fixed choice of $(V_{j+1}, \ldots, V_k)$ we have

$$\mathbb{E}\| *_j(\omega_{1,a_1}', \ldots, \omega_{j-1,a_{j-1}}', \nu_{j+1}, \ldots, \nu_k)\|_1 \;=\; \| *_j(1, \ldots, 1, \nu_{j+1}, \ldots, \nu_k)\|_1$$
$$\leq\; (p/q)^{k-j}\| *_j(1, \ldots, 1, \zeta_{j+1}, \ldots, \zeta_k)\|_1,$$

where the expectation here is taken over all sequences $(a_1, \ldots, a_{j-1})$. The inequality follows from the fact that $0 \leq \nu_h \leq (p/q)\zeta_h$ for each i. The constant η is at most 1, so applying assumption (i) if $j = 1$, we find that

$$\| *_1(\zeta_2, \ldots, \zeta_k)\|_1 \leq \| \circ_1(\zeta_2, \ldots, \zeta_k)\|_1 + \| *_1(\zeta_2, \ldots, \zeta_k) - \circ_1(\zeta_2, \ldots, \zeta_k)\|_1 \leq 2 + \eta \leq 3.$$

Similarly, applying assumption (ii) if $j > 1$, we have $\| *_j(1, \ldots, 1, \zeta_{j+1}, \ldots, \zeta_k)\|_1 \leq 2$. In either case,

$$\mathbb{E}\| *_j(\omega_{1,a_1}', \ldots, \omega_{j-1,a_{j-1}}', \nu_{j+1}, \ldots, \nu_k)\|_1 \leq 3(p/q)^{k-1} \leq 3L^k.$$

As $|X|$ tends to infinity, the probability that the first bound does not hold for every $(a_1, \ldots, a_{j-1})$ tends to zero, and the second bound always holds. Therefore, if $|X|$ is sufficiently large, it follows that

$$
\begin{aligned}
\|H\|_1 &\leq 2\eta + \mathbb{E}\| *_j (\omega'_{1,a_1}, \ldots, \omega'_{j-1,a_{j-1}}, \nu_{j+1}, \ldots, \nu_k) - \circ_j(\omega'_{1,a_1}, \ldots, \omega'_{j-1,a_{j-1}}, \nu_{j+1}, \ldots, \nu_k)\|_1 \\
&\leq 4\eta = \alpha,
\end{aligned}
$$

where the expectation is taken over all sequences containing those $\nu_{j+1}, \ldots, \nu_k$ such that, for all choices of $a_1, \ldots, a_{j-1}$, $(\omega_{1,a_1}, \ldots, \omega_{j-1,a_{j-1}}, \nu_{j+1}, \ldots, \nu_k)$ satisfies Q1. The result follows. $\qquad\square$

What we have shown is not just that every element of $\Phi(\zeta_{j+1}, \ldots, \zeta_k)$ can be approximated well in L_1 and reasonably well in L_∞ by a convex combination of elements of $\Psi(\zeta_{j+1}, \ldots, \zeta_k)$, but rather the stronger statement that the difference is bounded above by a *fixed* bounded function with small L_1-norm. This will be important to us later.

The title of this section was "The set of basic anti-uniform functions has few extreme points." That is an oversimplification: the next result is what we actually mean.

Lemma 6.11. *Let $0 < \alpha \leq 1/2k$ and $L \geq 2$ be a positive integer with $p = Lq$. Then, for $|X|$ sufficiently large depending on L and α, the following holds. Let $Z_{j+1}, \ldots, Z_k$ be subsets of X with associated measures $\zeta_{j+1}, \ldots, \zeta_k$ defined with weight p^{-1}, and suppose that assumptions (i), (ii) and (iii) are satisfied. Then there is a collection $\Psi' = \Psi'(\zeta_{j+1}, \ldots, \zeta_k)$ of at most $|X|^{k-1}\binom{2p|X|}{2q|X|}^{k-j}(2/\alpha)^{(k-1)2q|X|}$ functions that take values in $[0,2]$, and non-negative functions H and H' with $\|H\|_1 \leq \alpha$, $\|H\|_\infty \leq 2$, $\|H'\|_1 \leq 3\alpha(k-1)$ and $\|H'\|_\infty \leq 2$, such that for every function $\phi = \circ_j(g, \ldots, g, f_{j+1}, \ldots, f_k)$ in $\Phi(\zeta_{j+1}, \ldots, \zeta_k)$ there is a function ψ in the convex hull of Ψ' with $\psi \leq \phi \leq \psi + H + H'$.*

Proof. Choose a positive integer t such that $\alpha/2 \leq t^{-1} \leq \alpha$. Then there exists a non-negative function g' with $0 \leq g' \leq 1$ such that every value taken by g' is a multiple of t^{-1}, and such that $0 \leq g' \leq g \leq g' + \alpha$. Also, for every h and every function f_h such that $0 \leq f_h \leq \zeta_h$ there exists a function f'_h with $0 \leq f'_h \leq \zeta_h$ taking values that are multiples of $t^{-1}\zeta_h$ such that $0 \leq f'_h \leq f_h \leq f'_h + \alpha\zeta_h$.

We would now like to show, for any such choice of g and $f_{j+1}, \ldots, f_k$, that the functions $\phi = \circ_j(g, \ldots, g, f_{j+1}, \ldots, f_k)$ and $\phi' = \circ_j(g', \ldots, g', f'_{j+1}, \ldots, f'_k)$ are reasonably close. We shall consider the two cases $j = 1$ and $j > 1$ separately.

If $j = 1$ then

$$
\begin{aligned}
\phi - \phi' &= \sum_{h=2}^{k}(\circ_1(f'_2, \ldots, f'_{h-1}, f_h, \ldots, f_k) - \circ_1(f'_2, \ldots, f'_h, f_{h+1}, \ldots, f_k)) \\
&\leq \sum_{h=2}^{k}(*_1(f'_2, \ldots, f'_{h-1}, f_h, \ldots, f_k) - *_1(f'_2, \ldots, f'_h, f_{h+1}, \ldots, f_k)) \\
&= \sum_{h=2}^{k} *_1(f'_2, \ldots, f'_{h-1}, f_h - f'_h, f_{h+1}, \ldots, f_k) \\
&\leq \sum_{h=2}^{k} *_1(\zeta_2, \ldots, \zeta_{h-1}, \alpha\zeta_h, \zeta_{h+1}, \ldots, \zeta_k) \\
&= \alpha(k-1)(*_1(\zeta_2, \ldots, \zeta_k)).
\end{aligned}
$$

34

Since $\eta \leq 1$, assumption (i) implies that

$$\| *_1 (\zeta_2, \ldots, \zeta_k)\|_1 \leq 3,$$

so we find that $\circ_1(f_2, \ldots, f_k) - \circ_1(f'_2, \ldots, f'_k)$ is bounded above by a function H' with L_1-norm at most $3\alpha(k-1)$. It is clearly also bounded above by 2.

Lemma 6.10 gives us H with $\|H\|_1 \leq \alpha$, $\|H\|_\infty \leq 2$, and also $\psi \in \Psi(\zeta_2, \ldots, \zeta_k)$ such that $0 \leq \circ_1(f'_2, \ldots, f'_k) - \psi \leq H$. Putting these two facts together implies the required bounds on H and H' for the case $j = 1$.

If $j > 1$ then a very similar argument shows that

$$\phi - \phi' \quad \leq \quad \alpha(k-1)(*_j(1, \ldots, 1, \zeta_{j+1}, \ldots, \zeta_k)).$$

By assumption (ii), $\| *_j (1, \ldots, 1, \zeta_{j+1}, \ldots, \zeta_k)\|_\infty \leq 2$, so in this case we have a function H' with L_∞-norm at most $2\alpha(k-1) \leq 2$ and therefore with L_1-norm at most $2\alpha(k-1)$.

All that remains is to count the number of functions in $\Psi(\zeta_{j+1}, \ldots, \zeta_k)$ that are normalized restrictions of functions of the form $\circ_j(g', \ldots, g', f'_{j+1}, \ldots, f'_k)$. It is here that we shall use the assumption that the sets Z_h each have cardinality $(1 + o(1))p|X| \leq 2p|X|$. There are at most $|X|^{j-1}$ choices for the set $(a_1, \ldots, a_{j-1})$, and for each $j + 1 \leq i \leq k$, because of the upper bound on the sizes of the Z_i and V_i, there are at most $|X|\binom{2p|X|}{2q|X|}$ choices for the set V_i. (Note that since $p = Lq \geq 2q$, the largest binomial coefficient is indeed this one.) Finally, each valuation of each function has at most $t \leq 2/\alpha$ possible results and each of the $k - 1$ functions has a domain of size at most $2q|X|$. Therefore, the number of normalized restrictions is at most

$$|X|^{k-1} \left(\frac{2p|X|}{2q|X|}\right)^{k-j} (2/\alpha)^{(k-1)2q|X|},$$

as required. $\qquad\qquad\square$

6.3   The proof for products of basic anti-uniform functions

To connect the results of the previous subsection with basic anti-uniform functions, take a sequence $U_1, \ldots, U_m$ of subsets of X with associated measures $\mu_1, \ldots, \mu_m$. Then, for each j and each sequence $(i_{j+1}, \ldots, i_k)$ of distinct indices between 1 and m, we shall apply the results with $Z_h = U_{i_h}$ and $\zeta_h = \mu_{i_h}$. Then the functions in the set $\Phi(\zeta_{j+1}, \ldots, \zeta_k)$ are basic anti-uniform functions.

In this section, it will be clear that we are talking about measures $\mu_1, \ldots, \mu_m$, and therefore it will be convenient to write $\Phi(i_{j+1}, \ldots, i_k)$ and $\Psi(i_{j+1}, \ldots, i_k)$ instead of $\Phi(\mu_{i_{j+1}}, \ldots, \mu_{i_k})$ and $\Psi(\mu_{i_{j+1}}, \ldots, \mu_{i_k})$.

Our next task is to generalize Lemma 6.11 to a result that applies not just to basic anti-uniform functions but also to products of at most d such functions. This is a formal consequence of Lemma 6.11. The exact nature of the bounds we obtain for $\|J\|_1$ and $\|J\|_\infty$ is unimportant: what matters is that the first can be made arbitrarily small and the second is bounded. We need a definition.

Definition 6.12. *If $\phi \in \Phi(i_{j+1}, \ldots, i_k)$, then define the* profile *of ϕ to be the ordered set $(i_{j+1}, \ldots, i_k)$, and if ξ is a product of d basic anti-uniform functions ϕ_h, then define the profile of ξ to be the set of all d profiles of the ϕ_h. We will refer to d as the* size of the profile.

Corollary 6.13. *Let $0 < \alpha \le 1/2k$ and $L \ge 2$ be a positive integer with $p = Lq$. Then, for $|X|$ sufficiently large depending on L and α, the following holds. Suppose that A is a profile of size d and, for every $(i_{j+1}, \ldots, i_k)$ in A, the sets $U_{i_{j+1}}, \ldots, U_{i_k}$ satisfy assumptions (i), (ii) and (iii). Then there is a collection $\Delta = \Delta(A)$ of at most $|X|^{kd} \binom{2p|X|}{2q|X|}^{kd} (2/\alpha)^{2kdq|X|}$ functions that take values in $[0, 2^d]$ and a non-negative function $J = J(A)$ with $\|J\|_1 \le d\alpha k 6^d$ and $\|J\|_\infty \le d6^d$, such that for every function ξ that is a product of basic anti-uniform functions with profile A, there is a function ψ in the convex hull of Δ with $\psi \le \xi \le \psi + J$.*

Proof. Every function ξ with profile A is a product $\phi_1 \ldots \phi_d$, where each ϕ_i is a basic anti-uniform function with some fixed profile. That is, each ϕ_i belongs to a fixed set of the form $\Phi(i_{j+1}, \ldots, i_k)$. By Lemma 6.11 we can find ψ_i such that $\psi_i \le \phi_i \le \psi_i + J_i$, where ψ_i belongs to the convex hull of a set Ψ'_i of size at most $|X|^k \binom{2p|X|}{2q|X|}^k (2/\alpha)^{2kq|X|}$, and J_i is a fixed function such that $\|J_i\|_\infty \le 4$ and $\|J_i\|_1 \le 4\alpha k$.

It follows that $\prod_i \psi_i \le \xi \le \prod_i (\psi_i + J_i)$. But

$$\prod_{i=1}^d (\psi_i + J_i) - \prod_{i=1}^d \psi_i \le \sum_{h=1}^d J_h \prod_{i \ne h} (\psi_i + J_i).$$

Since each ψ_i has L_∞-norm at most 2, the latter function has L_1-norm at most $d\alpha k 6^d$ and L_∞-norm at most $d6^d$, as claimed. $\qquad\square$

We are now ready for the main result of this section. It will be convenient once again to give names to certain assumptions.

R1(r, j). If $Z_1, \ldots, Z_k$ are chosen independently from X_r and their associated measures are $\zeta_1, \ldots, \zeta_k$, then
$\| *_j (\zeta_1, \ldots, \zeta_{j-1}, \zeta_{j+1}, \ldots, \zeta_k) - \circ_j (\zeta_1, \ldots, \zeta_{j-1}, \zeta_{j+1}, \ldots, \zeta_k) \|_1 \le \eta$ with probability $1 - o(|X|^{-k})$.

R2(r). With the notation as in R1, the probability that $\| *_j (1, \ldots, 1, \zeta_{j+1}, \ldots, \zeta_k) \|_\infty \le 2$ for every $j \ge 2$ is $1 - o(1)$.

Note that R1(r, j) is saying that Q1 holds with high probability, and R2(r) is saying that Q2 holds with high probability for every j (when the ν_i are the associated measures of random sets from X_r).

Lemma 6.14. *For any positive constant λ and positive integer d, there exist η, m and L such that the following holds. Let $0 < p_0 \le 1/L$ and suppose that assumptions R1(r, j) and R2(r) hold for every j and for every $r \ge p_0$. Let $p \ge Lp_0$, let $U_1, \ldots, U_m$ be chosen independently from X_p, and let $\mu_1, \ldots, \mu_m$ be their associated measures. Then, with probability $1 - o(1)$, they satisfy property P3. That is, setting $\mu = m^{-1}(\mu_1 + \cdots + \mu_m)$, $|\langle \mu - 1, \xi \rangle| < \lambda$ whenever ξ is a product of at most d basic anti-uniform functions from $\Phi_{\mu,1}$.*

Proof. Let A be a profile and suppose that i is not involved in A. Let $\Gamma = \Gamma(A)$ be the set of all products of at most d basic anti-uniform functions with profile A. Let $q = p/L$ for a constant L yet to be determined. By assumption R1(q, j), Lemma 6.8 implies that for every sequence $(i_{j+1}, \ldots, i_k)$, the probability that assumption (iii) holds with parameters p and q for the sets $U_{i_{j+1}}, \ldots, U_{i_k}$ is $1 - o(1)$. Therefore, the probability that it holds for all j and all sequences $(i_{j+1}, \ldots, i_k)$ in the profile A is also

$1 - o(1)$. By assumptions R1$(p, 1)$ and R2(p), we also know that assumptions (i) and (ii) hold with probability $1 - o(1)$ for any given $(i_{j+1}, \ldots, i_k)$ and, therefore, for all $(i_{j+1}, \ldots, i_k)$ in the profile A.

We may therefore apply Corollary 6.13 to conclude that there exists a set $\Delta = \Delta(A)$ of at most $|X|^{kd} \binom{2p|X|}{2q|X|}^{kd} (2/\alpha)^{2kdq|X|}$ functions such that, for every function $\xi \in \Gamma$, there exists ψ in Δ with $|\xi - \psi| \leq H$, where $\|H\|_1 \leq d\alpha k6^d$ and $\|H\|_\infty \leq d6^d$. If we let $\alpha = \lambda/12kd6^d$, Corollary 6.3 implies that, with probability $1 - o(1)$,

$$\max\{|\langle \mu_i - 1, \phi \rangle| : \phi \in \Gamma\} \leq \max\{|\langle \mu_i - 1, \psi' \rangle| : \psi' \in \Delta\} + \frac{\lambda}{4}.$$

Note that this step depends critically on the fact that μ_i is entirely independent of the set $\Delta(A)$. It was for this purpose that we chose m random sets $U_1, \ldots, U_m$ rather than one single random set U. This observation is also important in the next step, which is to prove that $\max\{|\langle \mu_i - 1, \psi' \rangle| : \psi' \in \Delta\} \leq \lambda/4$ with probability $1 - o(1)$.

By Lemma 5.2, since $\|\psi'\|_\infty \leq 2^d$ for all $\psi' \in \Delta$, the probability that $|\langle \mu_i - 1, \psi' \rangle| > \lambda/4$ is at most $2\exp(-\lambda^2 p|X|/2^{2d+10})$ for any given ψ'. Since $p = Lq$, we may estimate the number of elements in $\Delta(A)$ as follows.

$$
\begin{aligned}
|X|^{kd} \binom{2p|X|}{2q|X|}^{kd} (2/\alpha)^{2kdq|X|} &\leq |X|^{kd} \left(\frac{6p|X|}{2q|X|}\right)^{2kdq|X|} \left(\frac{24kd6^d}{\lambda}\right)^{2kdq|X|} \\
&\leq |X|^{kd} (3L)^{2kdp|X|/L} \left(\frac{24kd6^d}{\lambda}\right)^{2kdp|X|/L} \\
&= |X|^{kd} \left(\left(\frac{72Lkd6^d}{\lambda}\right)^{2kd}\right)^{p|X|/L}.
\end{aligned}
$$

If we choose L sufficiently large (depending on k, d and λ), then we can arrange for the sum of the probabilities, which is at most $2\exp(-\lambda^2 p|X|/2^{2d+10})|\Delta(A)|$, to be $o(1)$.

We are almost done. We now wish to prove a result about $\mu = m^{-1}(\mu_1 + \cdots + \mu_m)$. Applying our result so far to all profiles simultaneously, we find that with probability $1 - o(1)$, $|\langle \mu_i - 1, \xi \rangle| \leq \lambda/2$ for every μ_i and ξ such that i is not involved in the profile of ξ. Fix a particular ξ_0. If we choose i at random, the probability that it is involved in the profile of ξ_0 is at most $(k-1)d/m$. Furthermore, for any i, we have the trivial bound $|\langle \mu_i - 1, \xi_0 \rangle| \leq 2^{d+2}$, since $\|\xi_0\|_\infty \leq 2^d$ and, for $|X|$ sufficiently large, $\|\mu_i - 1\|_1 \leq 3$. Therefore,

$$|\langle \mu - 1, \xi_0 \rangle| \leq \mathbb{E}_i |\langle \mu_i - 1, \xi_0 \rangle| \leq \frac{(k-1)d}{m} 2^{d+2} + \frac{\lambda}{2} \leq \lambda,$$

provided $m \geq kd2^{d+3}/\lambda$. The result follows. $\qquad\qquad\square$

6.4 Obtaining P3$'$ as well

It is possible to add a fixed set of bounded functions $\mathcal{F}$ to the collection of basic anti-uniform functions, provided only that this set has size smaller than $2^{p|X|/L_0}$, where L_0 is again some constant depending only on k, λ and d, and the above proof continues to work. Indeed, adding such a collection can increase the size of the set of products of basic anti-uniform functions by a factor of at most $2^{dp|X|/L_0}$.

Therefore, when we come to the final line of the penultimate paragraph of the proof of the previous lemma, provided L_0 and L have been chosen small enough, the probability that the random measure μ_i correlates with any given function is still small enough to guarantee that with high probability $\max\{|\langle \mu_i - 1, \psi'\rangle| : \psi' \in \Gamma'\} \leq \lambda/4$, where Γ' is the set of functions formed from products of at most d characteristic functions from $\mathcal{F}$ and basic anti-uniform functions whose profile does not involve μ_i. The remainder of the proof is the same, in that we add over all profiles and rule out the set of small exceptions where the set U_i is involved in the profile of ξ.

Later, when we come to apply this observation, $\mathcal{F}$ will be a collection of characteristic functions. For example, to prove a stability version of Turán's theorem, the set $\mathcal{F}$ will be the collection of characteristic measures of vertex subsets of $\{1, \ldots, n\}$. This has size 2^n. Therefore, provided $p \geq Cn^{-1}$, for C sufficiently large, we will have control over local densities.

7 Probabilistic estimates I: tail estimates

In this section, we shall focus on showing that property P2 holds with high probability. That is, we shall show that under suitable conditions, with high probability $\| *_j (1, 1, \ldots, 1, \mu_{i_{j+1}}, \ldots, \mu_{i_k})\|_\infty \leq 2$ for every $j \geq 2$ and every sequence $i_{j+1}, \ldots, i_k$ of distinct integers between 1 and m. It will be helpful for the next section if we actually prove the following very slightly more general statement. For every $1 \leq j \leq k$, every collection of measures $\nu_1, \ldots, \nu_k$ such that at least one of the measures other than ν_j is the constant measure 1 and the rest are distinct measures of the form μ_{i_j} has the property that $\| *_j (\nu_1, \ldots, \nu_k)\|_\infty \leq \frac{3}{2}$.

Up to now, our argument has been general. Unfortunately, we must now be more specific about the kind of sets that we are dealing with. We shall split into two cases. First, we shall look at systems S with the following property.

Definition 7.1. *A system S of ordered sequences of length k in a set X has two degrees of freedom if, whenever s and t are two elements of S and there exist $i \neq j$ such that $s_i = t_i$ and $s_j = t_j$, we have $s = t$.*

This includes the case when S is the set of arithmetic progressions in $\mathbb{Z}_n$, and higher-dimensional generalizations concerning homothetic copies of a single set.

After that, we will look at graphs and hypergraphs. In this case, the required estimates are much more difficult. Thankfully, most of the hard work has already been done for us by Janson, Ruciński and, in one paper, Oleszkiewicz [28, 29, 30, 31, 32] (see also the paper of Vu, [65]). We shall return to these estimates later.

7.1 The proof for systems with two degrees of freedom

Let $U_1, \ldots, U_m$ be independent random sets chosen binomially and let their associated measures be $\mu_1, \ldots, \mu_m$. We are interested in quantities of the form $*_j(\nu_1, \ldots, \nu_k)(x)$, where each ν_i (with $i \neq j$) is equal to either the constant function 1 or to one of the measures μ_r. We also insist that no two of the ν_i are equal to the same μ_r and that at least one of the ν_i is the constant function.

Suppose that the set of i such that ν_i is one of the μ_r is $\{a_1, \ldots, a_l\}$ and that $\nu_{a_h} = \mu_{b_h}$ for $h = 1, 2, \ldots, l$. Then we can interpret $*_j(\nu_1, \ldots, \nu_k)(x)$ as follows. Recall that $S_j(x)$ is the set of all $s = (s_1, \ldots, s_k) \in S$ such that $s_j = x$. Then $*_j(\nu_1, \ldots, \nu_k)(x)$ is equal to p^{-l} times the proportion of

$s \in S_j(x)$ such that $s_{a_h} \in U_{b_h}$ for every $h = 1, \ldots, l$. This is because $\nu_{a_h}(s_{a_h}) = p^{-1}$ if $s_{a_h} \in U_{b_h}$ and 0 otherwise.

Now let us regard sequences $s \in S$ as fixed and $U_1, \ldots, U_m$ as random variables. For each s, let $E(s)$ be the event that $s_{a_h} \in U_{b_h}$ for every $h = 1, \ldots, l$ (so $E(s)$ is an event that depends on $U_1, \ldots, U_m$). We claim that if s and t are distinct sequences in $S_j(x)$, then $E(s)$ and $E(t)$ are independent. The reason for this is that we know that $s_j = t_j$, and our assumption that S has two degrees of freedom therefore implies that there is no other i such that $s_i = t_i$. It follows that the events $s_{a_h} \in U_{b_h}$ and $t_{a_h} \in U_{b_h}$ are independent (since the sets U_i are chosen binomially) and hence that $E(s)$ and $E(t)$ are independent (since the sets $U_{b_1}, \ldots, U_{b_l}$ are independent).

Lemma 7.2. *Let X be a finite set, let S be a homogeneous collection of ordered subsets of X, each of size k, and suppose that S has two degrees of freedom. Let $U_1, \ldots, U_k$ be random subsets of X with associated measures $\mu_1, \ldots, \mu_k$, each chosen binomially with probability p. Let $1 \le j \le k$ and let L be a subset of $\{1, 2, \ldots, k\} \setminus \{j\}$ of cardinality $l < k - 1$. For each $i \le k$, let $\nu_i = \mu_i$ if $i \in L$ and 1 otherwise. Let $x \in \mathbb{Z}_n$. Then the probability that $*_j(\nu_1, \ldots, \nu_{j-1}, \nu_{j+1}, \ldots, \nu_k)(x) \le \frac{3}{2}$ is at least $1 - 2\exp(-p^l|S_j(x)|/16)$.*

Proof. Let χ_i be the characteristic function of U_i. Suppose that $L = \{a_1, \ldots, a_l\}$. Then

$$*_j(\nu_1, \ldots, \nu_{j-1}, \nu_{j+1}, \ldots, \nu_k)(x) = p^{-l}\mathbb{E}_{s \in S_j(x)} \prod_{i \in L} \chi_i(s_i).$$

Now $\prod_{i \in L} \chi_i(s_i)$ is the characteristic function of the event $E(s)$ mentioned just before the statement of this lemma, in the case when $b_h = a_h$ for every h. As we have discussed, these events are independent. Moreover, they each have probability p^l. Therefore, $\mathbb{E}_{s \in S_j(x)} \prod_{i \in L} \chi_i(s_i)$ is an average of $|S_j(x)|$ independent Bernoulli random variables of probability p^l.

By Chernoff's inequality, Lemma 5.3, $\sum_{s \in S_j(x)} \prod_{i \in L} \chi_i(s_i) \le \frac{3}{2}p^l|S_j(x)|$ with probability at least $1 - 2\exp(-p^l|S_j(x)|/16)$. Therefore, $\mathbb{E}_{s \in S_j(x)} \prod_{i \in L} \chi_i(s_i) \le \frac{3}{2}p^l$ with the same probability. This proves the result. $\qquad\square$

It is perhaps not immediately obvious how the bound for the probability in the last lemma relates to sharp values for p in applications. To get a feel for this, consider the case when S is the set of k-term arithmetic progressions in $\mathbb{Z}_n$. Then $|S_j(x)| = n$ for every x and j as $|X| = n$. We want to be able to take p to be around $n^{-1/(k-1)}$. With this value, $\exp(-p^l|S_j(x)|/16)$, takes the form $\exp(-cn^{1-l/(k-1)})$. In the worst case, when $l = k - 2$, this works out to be $\exp(-cn^{1/(k-1)})$, which drops off faster than any power of n. If we took $l = k - 1$ then we would no longer have an interesting statement: that is why convolutions where every ν_i is equal to some μ_j must be treated in a different way.

7.2 The proof for strictly balanced graphs and hypergraphs

We now turn to the more difficult case of finding copies of a fixed balanced graph or hypergraph. Again, we are trying to show that $*_j(\nu_1, \ldots, \nu_k)(x)$ is reasonably close to 1 with very high probability, but now this quantity is a normalized count of certain graphs or hypergraphs. Normally when one has a large deviation inequality, one expects the probability of large deviations to be exponentially small in the expectation. In the graph case a theorem of roughly this variety may be proved for the lower tail by using Janson's inequality [28], but the behaviour of the upper tail is much more complex. The

best that can be achieved is a fixed power of the expectation. The result that we shall use in this case is due to Janson and Ruciński [31]. Before we state it, we need some preliminary discussion.

To begin with, let us be precise about what we are taking as X and what we are taking as S. We are counting copies of a fixed labelled r-uniform hypergraph H. Let H have vertex set V of size m and (labelled) edge set $(e_1, \ldots, e_r)$. (That is, each e_i is a subset of V of size r and we choose some arbitrary ordering.) Let W be a set of size n (which we think of as large) and let $X = W^{(r)}$, the set of all subsets of W of order r.

Given any injection $\phi : V \to W$ we can form a sequence $(s_1, \ldots, s_k)$ of subsets of W by setting $s_i = \phi(e_i)$. We let S be the set of all sequences that arise in this way. The elements of S are copies of H with correspondingly labelled edges.

If we fix an edge $e \in X$ and an index j, then $S_j(e)$ is the set of all sequences $(s_1, \ldots, s_k)$ in S such that $s_j = e$. To obtain such a sequence, one must take a bijection from e_j (which is a subset of V of order r) to e (which is a subset of W of order r) and extend it to an injection ϕ from V to W. One then sets $s_i = \phi(e_i)$ for each i.

Now let $U_1, \ldots, U_m$ be independent random subsets of X, chosen binomially with probability p, and let their associated measures be $\mu_1, \ldots, \mu_m$. Suppose once again that $\nu_1, \ldots, \nu_k$ are measures, some of which are constant and some of which are equal to distinct μ_i. Suppose that the non-trivial measures, not including ν_j if it is non-trivial, are $\nu_{a_1}, \ldots, \nu_{a_l}$, and suppose that $\nu_{a_i} = \mu_{b_i}$ for $i = 1, 2, \ldots, l$. Then the value $*_j(\nu_1, \ldots, \nu_k)(e)$ of the jth convolution at e is equal to

$$\mathbb{E}_{s \in S_j(e)} \prod_{1 \leq i \leq l} \mu_{b_i}(s_{a_i}).$$

This is $p^{-l}|S_j(e)|^{-1}$ times the number of sequences $(s_1, \ldots, s_k) \in S$ such that $s_j = e$ and $s_{a_i} \in U_{b_i}$ for every $1 \leq i \leq l$. If we define H' to be the subhypergraph of H that consists of the edges $e_{a_1}, \ldots, e_{a_l}$, then each such sequence is a so-called e_j-*rooted* copy of H' in (e, X). That is, it is a copy of H' where we insist that the vertices in e_j map bijectively to the vertices in e. We are interested in the number of rooted copies such that the edges fall into certain sparse random sets. This is not an easy calculation, but it has been done for us by Janson and Ruciński. In order to state the result we shall need, let us define formally the random variable that we wish not to deviate much from its mean.

Notation. *Let K be a labelled r-uniform hypergraph and f an edge in K. Let l be the number of edges in $K \backslash \{f\}$ and let $U_1, \ldots, U_l$ be random binomial subhypergraphs of the complete r-uniform hypergraph $K_n^{(r)}$ on n vertices, each edge being chosen with probability p, with characteristic functions $\chi_1, \ldots, \chi_l$. Let S_f be the set consisting of all labelled ordered copies of $K \backslash \{f\}$ in $K_n^{(r)}$ that are f-rooted at a given edge e. Then the random variable Y_K^f is given by*

$$\sum_{s \in S_f} \prod_{1 \leq i \leq l} \chi_i(s_i).$$

Strictly speaking Y_K^f depends on e as well, but we omit this from the notation because it makes no difference to the probabilities which edge e we choose. (So we could, for example, state the result for $e = \{1, 2, \ldots, r\}$ and deduce it for all other e.)

The number of injections ϕ that extend a bijection from f to e is $r!(n-r)(n-r-1) \ldots (n-v_K+1)$, and for each one the probability that $s_i \in U_i$ for every i is $p^l = p^{e_K-1}$, so the expectation $\mathbb{E}Y_K^f$ is

$$p^{e_K-1}r!(n-r)(n-r-1) \ldots (n-v_K+1).$$

The precise details will not matter to us much, but note that the order of magnitude is $p^{e_K-1}n^{v_K-r}$.

We are now ready to state the result of Janson and Ruciński. It is actually a very special case of a much more general result (Corollary 4.1 from [31]). To explain the general statement would lead us too far astray so we restrict ourselves to stating the required corollary.

Lemma 7.3. *Let K be a labelled r-uniform hypergraph and f a fixed edge. Then there exists a constant c such that the random variable Y_K^f satisfies*

$$\mathbb{P}\left(Y_K^f \geq \frac{3}{2}\mathbb{E}Y_K^f\right) \leq 2n^{v_K}\exp\left(-c\min_{L\subseteq K}(\mathbb{E}Y_L^f)^{1/v_L}\right).$$

A better, indeed almost sharp, result has recently been proved by Janson and Ruciński [32]. Unfortunately, though the result almost certainly extends to hypergraphs, it is stated by these authors only for graphs. However, the previous result is more than sufficient for our current purposes.

We are now ready to show that if $X = K_n^{(r)}$, S is the collection of labelled copies of a strictly balanced hypergraph H in X and $p \geq n^{-1/m_r(H)}$, then P2 holds with high probability. The proof is essentially the same as it was for systems with two degrees of freedom, except that we have to use the results of Janson and Ruciński instead of Chernoff's inequality. Recall that an r-uniform hypergraph H is strictly r-balanced if $\frac{e_H-1}{v_H-r} > \frac{e_K-1}{v_K-r}$ for every proper subhypergraph K of H.

Lemma 7.4. *Let H be a strictly r-balanced r-uniform hypergraph with k edges. Let $X = K_n^{(r)}$ and let S be the collection of labelled ordered copies of H in X. Let $U_1, \ldots, U_k$ be random subsets of X, each chosen binomially with probability p, and let their characteristic measures be $\mu_1, \ldots, \mu_k$. Let $1 \leq j \leq k$ and let L be a subset of $\{1, 2, \ldots, k\} \setminus \{j\}$ of cardinality $l < k - 1$. For each $i \leq k$, let $\nu_i = \mu_i$ if $i \in L$ and 1 otherwise. Let $e \in X$. Then for $p \geq n^{-1/m_r(H)}$ there exist positive constants a and A such that the probability that $*_j(\nu_1, \ldots, \nu_{j-1}, \nu_{j+1}, \ldots, \nu_k)(e) \leq \frac{3}{2}$ is at least $1 - 2n^{v_H}e^{-An^a}$.*

Proof. Let χ_i be the characteristic function of U_i. Then

$$*_j(\nu_1, \ldots, \nu_{j-1}, \nu_{j+1}, \ldots, \nu_k)(e) = p^{-l}\mathbb{E}_{s\in S_j(e)}\prod_{i\in L}\chi_i(s_i).$$

The sum $\sum_{s\in S_j(e)}\prod_{i\in L}\chi_i(s_i)$ counts the number of rooted copies of some proper subhypergraph K of H. By Lemma 7.3, the probability that $\sum_{s\in S_j(e)}\prod_{i\in L}\chi_i(s_i) \geq \frac{3}{2}p^l|S_j(e)|$ is at most

$$2n^{v_K}\exp\left(-c\min_{J\subseteq K}(\mathbb{E}Y_J^{e_j})^{1/v_J}\right) = 2n^{v_K}\exp\left(-c'\min_{J\subseteq K}(n^{v_J-r}p^{e_J-1})^{1/v_J}\right).$$

Since H is strictly r-balanced, we know that $\frac{e_H-1}{v_H-r} > \frac{e_J-1}{v_J-r}$ for every $J \subseteq K$. Therefore, there is a positive constant a' such that if $p \geq n^{-1/m_r(H)}$, then for each $J \subseteq K$ we have the inequality

$$n^{v_J-r}p^{e_J-1} \geq n^{v_J-r}n^{-\left(\frac{v_H-r}{e_H-1}\right)(e_J-1)} \geq \left(n^{1-\left(\frac{v_H-r}{e_H-1}\right)\left(\frac{e_J-1}{v_J-r}\right)}\right)^{v_J-r} \geq n^{a'}.$$

Therefore,

$$\min_{J\subseteq K}(n^{v_J-r}p^{e_J-1})^{1/v_J} \geq n^a,$$

for some a, and hence the probability that $\sum_{s\in S_j(e)}\prod_{i\in L}\chi_i(s_i) \geq \frac{3}{2}p^l|S_j(e)|$ is at most $2n^{v_H}e^{-An^a}$ for some positive constants A and a. The lemma follows. $\qquad\square$

8 Probabilistic estimates II: bounding L_1-differences

Our one remaining task is to show that property P1 holds with sufficiently high probability. In other words, we must show that if $U_1, \ldots, U_m$ are subsets of X chosen binomially with suitable probability p, and if their associated measures are $\mu_1, \ldots, \mu_m$, then with high probability

$$\| *_j (\mu_{i_1}, \ldots, \mu_{i_{j-1}}, \mu_{i_{j+1}}, \ldots, \mu_{i_k}) - \circ_j (\mu_{i_1}, \ldots, \mu_{i_{j-1}}, \mu_{i_{j+1}}, \ldots, \mu_{i_k}) \|_1 \leq \eta$$

whenever j is an integer between 1 and k and $i_1, \ldots, i_{j-1}, i_{j+1}, \ldots, i_k$ are distinct integers between 1 and m. Of course, if we can prove this for one choice of j and $i_1, \ldots, i_{j-1}, i_{j+1}, \ldots, i_k$ then we have proved it for all, since m and k are bounded. So without loss of generality let us prove it for $j = 1$ and for the sequence $(2, \ldots, k)$. That is, we shall prove that with high probability

$$\| *_1 (\mu_2, \ldots, \mu_k) - \circ_1 (\mu_2, \ldots, \mu_k) \|_1 \leq \eta.$$

Our results will also imply the stronger statement R1$(p, 1)$, which was required for Lemma 6.14.

The basic approach is to show that with high probability the sets $U_2, \ldots, U_{k-1}$ have certain properties that we can exploit, and that if they have those properties then the conditional probability that $\| *_1 (\mu_2, \ldots, \mu_k) - \circ_1 (\mu_2, \ldots, \mu_k) \|_1 \leq \eta$ is also high. This strategy is almost forced on us: there are some choices of $U_2, \ldots, U_{k-1}$ that would be disastrous, and although they are rare we have to take account of their existence.

To get some idea of what the useful properties are, let us suppose that we have chosen $U_2, \ldots, U_{k-1}$, let us fix $x \in X$, and let us think about the random variable $*_1(\mu_2, \ldots, \mu_k)(x)$ (which, given our choices, depends just on the random set U_k). This is, by definition,

$$\mathbb{E}_{s \in S_1(x)} \mu_2(s_2) \ldots \mu_{k-1}(s_{k-1}) \mu_k(s_k).$$

At this point we need an extra homogeneity assumption. We would like to split up the above expectation according to the value of s_k, but that will lead to problems if different values of s_k are taken different numbers of times. Let us suppose that for each y the number of $s \in S_1(x)$ such that $s_k = y$, which is just the cardinality of the set $S_1(x) \cap S_k(y)$, only ever takes one of two values, one of which is 0.

In the case of arithmetic progressions of length k in $\mathbb{Z}_p$, with p prime, $S_1(x) \cap S_k(y)$ consists of a unique arithmetic progression (degenerate if $x = y$), the progression with common difference $(k-1)^{-1}(y-x)$ that starts at x. In the case of, say, K_5s in a complete graph, where s_1 and s_{10} represent disjoint edges of K_5, $S_1(e) \cap S_{10}(e')$ will be empty if e and e' are edges of K_n that share a vertex, and will have cardinality $n - 4$ if they are disjoint. In general, in all natural examples this homogeneity assumption is satisfied. Moreover, the proportion of y for which $S_1(x) \cap S_k(y) = \emptyset$ tends to be $O(1/n)$ and tends to correspond to degenerate cases (when those are not allowed).

With the help of this assumption, we can rewrite the previous expression as follows. Let us write $K(x)$ for the set of y such that $S_1(x) \cap S_k(y) \neq \emptyset$. Then

$$
\begin{aligned}
*_1(\mu_2, \ldots, \mu_k)(x) &= \mathbb{E}_{s \in S_1(x)} \mu_2(s_2) \ldots \mu_{k-1}(s_{k-1}) \mu_k(s_k) \\
&= \mathbb{E}_{y \in K(x)} \mu_k(y) \mathbb{E}_{s \in S_1(x) \cap S_k(y)} \mu_2(s_2) \ldots \mu_{k-1}(s_{k-1}).
\end{aligned}
$$

Writing $W(x, y)$ for $\mathbb{E}_{s \in S_1(x) \cap S_k(y)} \mu_2(s_2) \ldots \mu_{k-1}(s_{k-1})$, we can condense this to $\mathbb{E}_{y \in K(x)} \mu_k(y) W(x, y)$.

Now we are thinking of $\mu_2, \ldots, \mu_{k-1}$ as fixed, and of the expressions we write as random variables that depend on the random measure μ_k. Note that the expectation of $*_1(\mu_2, \ldots, \mu_k)(x)$ is $*_1(\mu_2, \ldots, \mu_{k-1}, 1)(x)$. By the results of the previous section, we are free to assume that this is at most $3/2$ for every x.

Our plan is to prove that the expectation of $*_1(\mu_2, \ldots, \mu_k)(x) - \circ_1(\mu_2, \ldots, \mu_k)(x)$ is small for each x, which will show that the expectation of $\| *_1 (\mu_2, \ldots, \mu_k) - \circ_1(\mu_2, \ldots, \mu_k)\|_1$ is small. Having done that, we shall argue that it is highly concentrated about its expectation.

Now, as we have seen, the random variable $*_1(\mu_2, \ldots, \mu_k)(x)$ is equal to $\mathbb{E}_{y \in K(x)} \mu_k(y) W(x, y)$, which is a sum of independent random variables V_y, where $V_y = (p|K(x)|)^{-1} W(x, y)$ with probability p and 0 otherwise. The expectation $\mathbb{E}_{y \in K(x)} W(x, y)$ of this sum is $*_1(\mu_2, \ldots, \mu_{k-1}, 1)(x)$, which we are assuming to be at most $3/2$. If we also know that each V_y is small, then the chances that this sum is bigger than 2 are very small. From this it is possible to deduce that the expectation of $*_1(\mu_2, \ldots, \mu_k)(x) - \circ_1(\mu_2, \ldots, \mu_k)(x)$ is small. The following technical lemma makes these arguments precise.

In the statement of the next lemma, we write $\mathbb{E}_{y \in K}$ for the average over K, and $\mathbb{E}$ for the probabilistic expectation (over all possible choices of μ_k with their appropriate probabilities).

Lemma 8.1. *Let $0 \leq p \leq 1$ and let $0 < \alpha \leq 1$. Let K be a set and for each $y \in K$ let V_y be a random variable that takes the value $C_y > 0$ with probability p and 0 otherwise. Suppose that the V_y are independent and that each C_y is at most α. Let $S = \sum_{y \in K} V_y$ and suppose that $\mathbb{E}S \leq 3/2$. Let $T = \max\{S - 2, 0\}$. Then $\mathbb{E}T \leq 14\alpha e^{-1/14\alpha}$.*

Proof. If we increase the number of random variables or any of the values C_y, then the expectation of T increases. Therefore, we are done if we can prove the result in the case where $\mathbb{E}S = 3/2$.

We shall use the elementary identity

$$\mathbb{E}T = \int_0^\infty \mathbb{P}[T \geq t]dt = \int_0^\infty \mathbb{P}[S \geq 2 + t]dt.$$

Since $\mathbb{E}S = 3/2$, if $S \geq 2 + t$ it follows that $S - \mathbb{E}S \geq t + 1/2$. Let us bound the probability of this event using Bernstein's inequality (Lemma 5.1).

For this we need to bound $\sum_y \mathbb{V}(V_y)$, which is at most $\sum_y \mathbb{E}(V_y^2)$, which is at most $\alpha \sum_y \mathbb{E}(V_y)$, by our assumption about the upper bound for each C_y. But this is $\alpha \mathbb{E}S = 3\alpha/2$. Therefore,

$$\mathbb{P}[S \geq 2 + t] \leq 2\exp\left\{\frac{-(t + 1/2)^2}{2(3\alpha/2 + \alpha(t + 1/2)/3)}\right\}.$$

Writing $s = t + 1/2$, this gives us $2\exp(-s^2/(3\alpha + 2\alpha s/3))$. When $s \geq 1/2$ (as it is everywhere in the integral we are trying to bound), this is at most $2\exp(-s^2/(6\alpha s + 2\alpha s/3)) \leq 2\exp(-s/7\alpha)$, so we have an upper bound of

$$2\int_{1/2}^\infty \exp(-s/7\alpha)ds = 14\alpha e^{-1/14\alpha},$$

which proves the lemma. $\square$

Corollary 8.2. *Suppose that $\mu_2, \ldots, \mu_{k-1}$ are fixed and that $W(x, y) \leq \alpha p|K(x)|$ for every x and y and $*_1(\mu_2, \ldots, \mu_{k-1}, 1)(x) \leq 3/2$ for every x. Then*

$$\mathbb{E}(*_1(\mu_2, \ldots, \mu_k)(x) - \circ_1(\mu_2, \ldots, \mu_k)(x)) \leq 14\alpha e^{-1/14\alpha}$$

for every x.

Proof. As noted above, $*_1(\mu_2, \ldots, \mu_k)(x)$ is a sum of independent random variables V_y that take the value $(p|K(x)|)^{-1}W(x,y)$ with probability p and 0 otherwise. By our hypothesis about $W(x,y)$, we can take $C_y = \alpha$ for each y and apply the previous lemma. Then $S = *_1(\mu_2, \ldots, \mu_k)(x)$ and $T = *_1(\mu_2, \ldots, \mu_k)(x) - \circ_1(\mu_2, \ldots, \mu_k)(x)$, so the result follows. $\qquad\square$

The next result but one is our main general lemma, after which we shall have to argue separately for different kinds of system. We shall use the following concentration of measure result, which is an easy and standard consequence of Azuma's inequality.

Lemma 8.3. *Let $X^{(t)}$ be the collection of all subsets of size t of a finite set X. Let $c, \lambda > 0$ and let F be a function defined on $X^{(t)}$ such that $|F(U) - F(V)| \leq c$ whenever $|U \cap V| = t - 1$. Then if a random set $U \in X^{(t)}$ is chosen, the probability that $|F(U) - \mathbb{E}F| \geq \lambda$ is at most $2\exp(-\lambda^2/2c^2 t)$.*

Most of the conditions of the next lemma have been mentioned in the discussion above, but we repeat them for convenience (even though the resulting statement becomes rather long).

Lemma 8.4. *Let X be a finite set and let S be a homogeneous collection of ordered subsets of X, each of size k. Let σ be a positive integer and suppose that, for all $x, y \in X$, $|S_1(x) \cap S_k(y)| \in \{0, \sigma\}$. For each x, let $K(x)$ be the set of y such that $S_1(x) \cap S_k(y) \neq \emptyset$, and suppose that all the sets $K(x)$ have the same size.*

Let $\mu_2, \ldots, \mu_{k-1}$ be fixed measures such that $*_1(\mu_2, \ldots, \mu_{k-1}, 1)(x)$ and $*_k(1, \mu_2, \ldots, \mu_{k-1})(x)$ are at most $3/2$ for every $x \in X$. For each $x, y \in X$, let

$$W(x,y) = \mathbb{E}_{s \in S_1(x) \cap S_k(y)} \mu_2(s_2) \ldots \mu_{k-1}(s_{k-1})$$

and suppose that $W(x,y) \leq \alpha p |K(x)|$ for every x and y.

Let U_k be a random set chosen binomially with probability p, let μ_k be its associated measure, and let $\eta = 28\alpha e^{-1/14\alpha}$. Then

$$\mathbb{P}[\| *_1(\mu_2, \ldots, \mu_k) - \circ_1(\mu_2, \ldots, \mu_k)\|_1 > \eta] \leq 2|X|e^{-\eta^2 p|X|/144} + 2e^{-p|X|/4}.$$

Proof. Corollary 8.2 and linearity of expectation imply that

$$\mathbb{E}\| *_1(\mu_2, \ldots, \mu_k) - \circ_1(\mu_2, \ldots, \mu_k)\|_1 \leq 14\alpha e^{-1/14\alpha} = \eta/2.$$

Let us write Z for the random variable $\| *_1(\mu_2, \ldots, \mu_k) - \circ_1(\mu_2, \ldots, \mu_k)\|_1$. To complete the proof, we shall show that Z is highly concentrated about its mean.

To do this, we condition on the size of the set U_k and apply Lemma 8.3. Suppose, then, that $|U_k| = t$. We must work out by how much we can change Z if we remove an element of U_k and add another.

Since the function $x \mapsto \max\{x - 2, 0\}$ is 1-Lipschitz, the amount by which we can change Z is at most the amount by which we can change $Y = \| *_1(\mu_2, \ldots, \mu_k)\|_1$. But

$$
\begin{aligned}
\| *_1(\mu_2, \ldots, \mu_k)\|_1 &= \mathbb{E}_x \mathbb{E}_{s \in S_1(x)} \mu_2(s_2) \ldots \mu_{k-1}(s_{k-1})\mu_k(s_k) \\
&= \mathbb{E}_y \mu_k(y)\mathbb{E}_{s \in S_k(y)} \mu_2(s_2) \ldots \mu_{k-1}(s_{k-1}) \\
&= \mathbb{E}_y \mu_k(y) *_k(1, \mu_2, \ldots, \mu_{k-1})(y).
\end{aligned}
$$

44

We are assuming that $*_k(1, \mu_2, \ldots, \mu_{k-1})(y)$ is never more than $3/2$, and $\mu_k(y)$ is always either p^{-1} or 0, so changing one element of U_k cannot change Y by more than $3(p|X|)^{-1}$. (The division by $|X|$ is because we are taking an average over y rather than a sum over y.)

Lemma 8.3 now tells us that the probability that $Z - \mathbb{E}Z \geq \eta/2$ given that $|U_k| = t$ is at most $2\exp(-\eta^2 p^2 |X|^2/72t)$. It follows that if $t \leq 2p|X|$ then the probability is at most $2\exp(-\eta^2 p|X|/144)$. By Chernoff's inequality, the probability that $t > 2p|X|$ is at most $2\exp(-p|X|/4)$. Putting these two facts together and adding over all possible values of t, we obtain the result stated. $\qquad\square$

Our aim is to prove that property P1 holds with high probability for a given small constant $\eta > 0$. Therefore, it remains to prove that, under suitable conditions on p, we have the bound $W(x, y) \leq \alpha p|K(x)|$ for every $x, y \in X$ such that $S_1(x) \cap S_k(y)$ is non-empty, where α is also a given small constant. Here, the argument once again depends on the particular form of the set of sequences S.

In the case of sets with two degrees of freedom, this is trivial. Let us suppose that $|K(x)| = t$ for every $x \in X$. By definition, $S_i(x) \cap S_j(y)$ is either empty or a singleton for every $1 \leq i < j \leq k$ and every $x, y \in X$. It follows, when $S_1(x) \cap S_k(y)$ is non-empty, that

$$
\begin{aligned}
W(x, y) &= \mathbb{E}_{s \in S_1(x) \cap S_k(y)} \mu_2(s_2) \ldots \mu_{k-1}(s_{k-1}) \\
&= \mu_2(r_2) \ldots \mu_{k-1}(r_{k-1}) \\
&\leq p^{-(k-2)},
\end{aligned}
$$

where $r = (x, r_2, \ldots, r_{k-1}, y)$ is the unique element of S that belongs to $S_1(x) \cap S_k(y)$. This is smaller than αpt as long as $p \geq (\alpha t)^{-1/(k-1)}$. Recall that in a typical instance, such as when S is the set of k-term arithmetic progressions in $\mathbb{Z}_n$ for some prime n, t will be very close to n (or in that case actually equal to n), and we do indeed obtain a bound of the form $Cn^{-1/(k-1)}$ that is within a constant of best possible.

Thus, we have essentially already finished the proof of a sparse random version of Szemerédi's theorem, and of several other similar theorems. We will spell out the details of these applications later in the paper. Now, however, let us turn to the more difficult task of verifying the hypothesis about W in the case of graphs and hypergraphs.

Let H be a strictly r-balanced r-uniform hypergraph. Recall that $m_r(H)$ is the ratio $(e_H - 1)/(v_H - r)$. The significance of $m_r(H)$ is that if $G_{n,p}^{(r)}$ is a random r-uniform hypergraph on n vertices, with each edge chosen with probability p, then the expected number of labelled copies of H containing any given edge of $G_{n,p}^{(r)}$ is approximately $p^{e_H-1}n^{v_H-r}$ (the "approximately" being the result of a few degenerate cases), so we need $p \geq n^{-1/m_r(H)}$ for this expected number to be at least 1, which, at least in the density case, is a trivial necessary condition for our theorems to hold. Our main aim now is to prove that $W(x, y) \leq \alpha p|K(x)|$ holds when $p \geq Cn^{-1/m_r(H)}$, where C is a constant that depends only on α and the hypergraph H.

In the next result, we shall take p to equal $Cn^{-1/m_r(H)}$ and prove that the conclusion holds provided C is sufficiently large. However, it turns out that we have to split the result into two cases. In the first case, we also need to assume that C is smaller than n^c for some small positive constant c, or else the argument breaks down. However, when C is larger than this (so not actually a constant) we can quote results of Janson and Ruciński to finish off the argument. (Some of our results, in particular colouring theorems, are monotone, in the sense that the result for p implies the result for all $q \geq p$. In such cases we do not need to worry about large p.)

Lemma 8.5. *Let H be a strictly r-balanced r-uniform hypergraph and let S be the collection of labelled ordered copies of H in the complete r-uniform hypergraph $K_n^{(r)}$. Then, for any positive constants α and A, there exist constants $c > 0$ and C_0 such that, if n is sufficiently large, $C_0 \le C \le n^c$, and $p = Cn^{-1/m_r(H)}$, then, with probability at least $1 - n^{-A}$, if $U_2, \ldots, U_{e-1}$ are random subgraphs $G_{n,p}^{(r)}$ of $K_n^{(r)}$ with associated measures $\mu_2, \ldots, \mu_{e-1}$,*

$$W(x, y) = \mathbb{E}_{s \in S_1(x) \cap S_e(y)} \mu_2(s_2) \ldots \mu_{e-1}(s_{e-1}) \le \alpha p |K(x)|,$$

for all $x, y \in X$, where we have written e for e_H.

Proof. Let χ_i be the characteristic function of U_i for each $i \le e_H$. Let σ be the size of each non-empty set $S_1(x) \cap S_e(y)$ and suppose $|K(x)| = t$ for each x. Then

$$W(x, y) = \sigma^{-1} p^{-(e_H - 2)} \sum_{s \in S_1(x) \cap S_e(y)} \chi_2(s_2) \ldots \chi_{e-1}(s_{e-1}).$$

But $\sum_{s \in S_1(x) \cap S_e(y)} \chi_2(s_2) \ldots \chi_{e-1}(s_{e-1})$ is the number of sequences $(s_1, \ldots, s_e) \in S$ such that $s_1 = x$, $s_e = y$ and $s_i \in U_i$ for $i = 2, 3, \ldots, e - 1$. Therefore, our aim is to prove that with high probability this number is at most $\alpha p t \sigma p^{e_H - 2} = \alpha p^{e_H - 1} \sigma t$. Let h be the number of vertices in the union of the first and eth edges. Then σ is almost exactly $n^{v_H - h}$ and t is almost exactly $n^{h - r}$, so it is enough to prove that with high probability the number of such sequences is at most $(\alpha/2) p^{e_H - 1} n^{v_H - r} = (\alpha/2) C^{e_H - 1}$. To do this, let us estimate from above the probability that there are at least $(v_H \ell)^{v_H}$ such sequences.

It will be convenient to think of each sequence in $S_1(x) \cap S_e(y)$ as an embedding ϕ from H to $K_n^{(r)}$ such that, writing $f_1, \ldots, f_e$ for the edges of H, we have $\phi(f_1) = x$ and $\phi(f_e) = y$. Let us call ϕ *good* if in addition $\phi(e_i) \in U_i$ for $i = 2, 3, \ldots, e - 1$. Now if there are $(v_H \ell)^{v_H}$ good embeddings, then there must be a sequence $\phi_1, \ldots, \phi_\ell$ of good embeddings such that each $\phi_i(H)$ contains at least one vertex that is not contained in any of $\phi_1(H), \ldots, \phi_{i-1}(H)$. That is because the number of vertices in the union of the images of the embeddings has to be at least $v_H \ell$, since the number of embeddings into a set of size u is certainly no more than u^{v_H}, and because each embedding has v_H vertices.

Let us fix a sequence of embeddings $\phi_1, \ldots, \phi_\ell$ such that each one has a vertex in its image that is not in the image of any previous one. Let $v_1, \ldots, v_m$ be the sequence of vertices obtained by listing all the vertices of $\phi_1(H)$ in order (taken from an initial fixed order of the vertices of H), then all the vertices of $\phi_2(H)$ that have not yet been listed, again in order, and so on. For each $i \le \ell$, let V_i be the set of vertices in $\phi_i(H)$ but no earlier $\phi_j(H)$. We shall now estimate the probability that every ϕ_i is good. If we already know that $\phi_1, \ldots, \phi_{i-1}$ are all good, then what we need to know is how many edges belong to $\phi_i(H)$ that do not belong to $\phi_j(H)$ for any $j < i$. Let $w_i = |V_i|$ be the number of vertices that belong to $\phi_i(H)$ and to no earlier $\phi_j(H)$, and let d_i be the number of edges. Then the conditional probability that ϕ_i is good is p^{d_i}. It follows that the probability that $\phi_1, \ldots, \phi_\ell$ are all good is $p^{d_1 + \cdots + d_\ell}$. The number of possible sequences of embeddings of this type is at most $m^{v_H \ell} n^m$, since there are at most n^m sequences $v_1, \ldots, v_m$, and once we have chosen $v_1, \ldots, v_m$ there are certainly no more than m^{v_H} ways of choosing the embedding ϕ_i (assuming that its image lies in the set $\{v_1, \ldots, v_m\}$). Therefore, the probability that there exists a good sequence of ℓ embeddings of this type is at most $m^{v_H \ell} p^{d_1 + \cdots + d_\ell} n^{w_1 + \cdots + w_\ell}$.

At this point, we use the hypothesis that H is strictly balanced. Since $w_i \le v_H - h \le v_H - (r + 1)$,

$$\frac{e_H - 1 - d_i}{v_H - r - w_i} < \frac{e_H - 1}{v_H - r},$$

46

which implies that $d_i/w_i > m_r(H)$. In fact, since there are only finitely many possibilities for w_i and d_i, it tells us that there is a constant $c' > 0$ depending on H only such that $d_i \geq m_r(H)(w_i + c')$. Since $p = Cn^{-1/m_r(H)}$, this tells us that $p^{d_i} \leq C^{d_i}n^{-(w_i+c')}$, and hence that

$$m^{v_H\ell}p^{d_1+\cdots+d_\ell}n^{w_1+\cdots+w_\ell} \leq m^{v_H\ell}C^{d_1+\cdots+d_\ell}n^{-\ell c'}.$$

To complete the proof, let us show how to choose C, just to be sure that the dependences are correct. We start by choosing ℓ such that $\ell c' \geq 2A$. Bearing in mind that $m \leq v_H\ell$ and that $d_1 + \cdots + d_\ell \leq e_H\ell$, we place on C the upper bound $C \leq (v_H\ell)^{-v_H\ell}n^{A/e_H\ell}$, which ensures that $m^{v_H\ell}C^{d_1+\cdots+d_\ell}n^{-\ell c'} \leq n^{-A}$. Finally, we need C to be large enough for $(\alpha/2)C^{e_H-1}$ to be greater than $(v_H\ell)^{v_H}$, since then the probability that there are at least $(\alpha/2)C^{e_H-1}$ sequences is at most n^{-A}, which is what we were trying to prove. Thus, we need C to be at least $(2(v_H\ell)^{v_H}/\alpha)^{1/(e_H-1)}$. $\qquad\square$

To handle the case where $C \geq n^c$, we shall again need to appeal to the work of Janson and Ruciński on upper tail estimates. The particular random variable we will be interested in, which concerns hypergraphs which are rooted on two edges, is defined as follows.

Notation. *Let K be an r-uniform hypergraph and f_1, f_2 edges in K. Let l be the number of edges in $K\backslash\{f_1, f_2\}$ and let $U_1, \ldots, U_l$ be random binomial subhypergraphs of the complete r-uniform hypergraph $K_n^{(r)}$ on n vertices, each edge being chosen with probability p, with characteristic functions $\chi_1, \ldots, \chi_l$. Let S_{f_1,f_2} be the set consisting of all labelled ordered copies of $K\backslash\{f_1, f_2\}$ in $K_n^{(r)}$ that are rooted at given edges e_1 and e_2. Then the random variable $Y_K^{f_1,f_2}$ is given by*

$$\sum_{s\in S_{f_1,f_2}} \prod_{1\leq i\leq l} \chi_i(s_i).$$

The necessary tail estimate (which is another particular case of Corollary 4.1 in [31]) is now the following. Note that $\mathbb{E}Y_K^{f_1,f_2}$ is essentially $p^{e_K-2}n^{v_K-h}$, where h is the size of $f_1 \cup f_2$.

Lemma 8.6. *Let K be an r-uniform hypergraph and f_1, f_2 fixed edges. Then there exists a constant c such that the random variable $Y_K^{f_1,f_2}$ satisfies, for $\gamma \geq 2$,*

$$\mathbb{P}\left(Y_K^{f_1,f_2} \geq \gamma\mathbb{E}Y_K^{f_1,f_2}\right) \leq 2n^{v_K} \exp\left(-c\min_{L\subseteq K}\left(\gamma\mathbb{E}Y_L^{f_1,f_2}\right)^{1/v_L}\right).$$

The required estimate for $p \geq n^{-1/m_k(H)+c}$ is now an easy consequence of this lemma.

Lemma 8.7. *Let H be a strictly r-balanced r-uniform hypergraph and let S be the collection of labelled ordered copies of H in the complete r-uniform hypergraph $K_n^{(r)}$. Then, for any positive constants α and c, there exist constants b and B such that, if n is sufficiently large, $C \geq n^c$, and $p = Cn^{-1/m_r(H)}$, then, with probability at least $1 - 2n^{v_H}e^{-Bn^b}$, if $U_2, \ldots, U_{e-1}$ are random subgraphs $G_{n,p}^{(r)}$ of $K_n^{(r)}$ with associated measures $\mu_2, \ldots, \mu_{e-1}$,*

$$W(x, y) = \mathbb{E}_{s\in S_1(x)\cap S_e(y)}\mu_2(s_2)\ldots\mu_{e-1}(s_{e-1}) \leq \alpha p|K(x)|,$$

for all $x, y \in X$, where we have written e for e_H.

Proof. Let χ_i be the characteristic function of U_i for each $i \leq e_H$. Let σ be the size of each non-empty set $S_1(x) \cap S_e(y)$ and suppose $|K(x)| = t$ for each x. Note that $\mathbb{E}Y_H^{e_1,e_H} = p^{e_H-2}\sigma$. We may apply Lemma 8.6 with $\gamma = \alpha pt$ to tell us that $\sum_{s \in S_1(x) \cap S_e(y)} \chi_2(s_2) \ldots \chi_{e-1}(s_{e-1}) \geq \gamma p^{e_H-2}\sigma$ with probability at most

$$2n^{v_H} \exp\left(-c \min_{L \subseteq H} \left(\gamma \mathbb{E}Y_L^{e_1,e_H}\right)^{1/v_L}\right) = 2n^{v_H} \exp\left(-c \min_{L \subseteq H} \left(\gamma n^{v_L-h}p^{e_L-2}\right)^{1/v_L}\right),$$

where h is the size of $e_1 \cup e_H$. Note that, as t is almost exactly n^{h-r}, $\gamma n^{v_L-h}p^{e_L-2} \geq (\alpha/2)n^{v_L-r}p^{e_L-1}$. Since H is strictly r-balanced, for any proper subgraph L of H,

$$n^{v_L-r}p^{e_L-1} \geq \left(n^{1-\left(\frac{v_H-r}{e_H-1}\right)\left(\frac{e_L-1}{v_L-r}\right)}\right)^{v_L-r} \geq n^{b'}.$$

Since also $n^{v_H-r}p^{e_H-1} \geq n^{\epsilon}$, the required bound holds with probability at least $1 - 2n^{v_H}e^{-Bn^b}$ for some constants B and b. Since

$$W(x,y) = \sigma^{-1}p^{-(e_H-2)} \sum_{s \in S_1(x) \cap S_e(y)} \chi_2(s_2) \ldots \chi_{e-1}(s_{e-1}),$$

the result now follows for n sufficiently large. □

9 Summary of our results so far

We are about to discuss several applications of our main results. In this brief section, we prepare for these applications by stating the abstract results that follow from the work we have done so far. Since not every problem one might wish to solve will give rise to a system of sequences S that either has two degrees of freedom or concerns copies of a strictly balanced graph or hypergraph, we begin by stating sufficient conditions on S for theorems of the kind we are interested in to hold. We have of course already done this, but since some of our earlier conditions implied other ones, there is scope for stating the abstract results more concisely. That way, any further applications of our methods will be reduced to establishing two easily stated probabilistic estimates, and showing that suitable robust versions of the desired results hold in the dense case.

Having done that, we remark that we have proved that the estimates hold when S has two degrees of freedom or results from copies of a strictly balanced graph or hypergraph. So in these two cases, if the robust results hold in the dense case, then we can carry them over unconditionally to the sparse random case.

The proofs in this section require little more than the putting together of results from earlier in the paper.

9.1 Conditional results

Recall that a system S of sequences $s = (s_1, \ldots, s_k)$ with values in a finite set X is *homogeneous* if for every $j \leq k$ and every $x \in X$ the set $S_j(x) = \{s \in S : s_j = x\}$ has the same size. Let S be a homogeneous system of sequences with elements in a finite set X, and let us assume that no sequence in S has repeated elements. We shall also assume that all non-empty sets of the form $S_1(x) \cap S_k(y)$

have the same size. Coupled with our first homogeneity assumption, this implies that for each x the number of y such that $S_1(x) \cap S_k(y)$ is non-empty is the same.

We are about to state and prove a theorem that is similar to Theorem 4.5, but with conditions that are easier to check and a conclusion that is more directly what we want to prove. The first condition is what we proved in Lemmas 7.2 and 7.4. We suppose that X is a given finite set, S is a given homogeneous system of sequences with terms in X, and p_0 is a given probability.

Condition 1. *Let* $U_1, \ldots, U_k$ *be independent random subsets of* X, *each chosen binomially with probability* $p \geq p_0$, *and let* $\mu_1, \ldots, \mu_k$ *be their associated measures. Let* $1 \leq j \leq k$ *and for each* $i \neq j$ *let* ν_i *equal either* μ_i *or the constant measure 1 on* X, *with at least one* ν_i *equal to the constant measure. Then with probability at least* $1 - o(|X|^{-k})$,

$$*_j(\nu_1, \ldots, \nu_{j-1}, \nu_{j+1}, \ldots, \nu_k)(x) \leq 3/2$$

for every $x \in X$.

Recall that if L is the set of i such that $\nu_i = \mu_i$, then $*_j(\nu_1, \ldots, \nu_{j-1}, \nu_{j+1}, \ldots, \nu_k)(x)$ is $p^{-|L|}$ times the number of $s \in S_j(x)$ such that $s_i \in U_i$ for every $i \in L$. Since the expected number of such sequences is $p^{|L|}|S_j(x)|$, Condition 1 is saying that their number is not too much larger than its mean. (One would usually expect a concentration result that said that their number is, with high probability, close to its mean.)

The second condition tells us that the hypotheses of Lemma 8.4 hold. Again we shall take X, S and p as given.

Condition 2. *Let* $U_2, \ldots, U_{k-1}$ *be independent random subsets of* X, *each chosen binomially with probability* $p \geq p_0$, *and let* $\mu_2, \ldots, \mu_{k-1}$ *be their associated measures. Let* $\alpha > 0$ *be an arbitrary positive constant. For each* x, *let* t *be the number of* y *such that* $S_1(x) \cap S_k(y)$ *is non-empty. Then with probability at least* $1 - o(|X|^{-k})$,

$$W(x, y) = \mathbb{E}_{s \in S_1(x) \cap S_k(y)} \mu_2(s_2) \ldots \mu_{k-1}(s_{k-1}) \leq \alpha p t$$

for every x, y *such that* $S_1(x) \cap S_k(y)$ *is non-empty.*

This is not a concentration assumption. For instance, in the case of systems with two degrees of freedom, it follows trivially from the fact that $|S_1(x) \cap S_k(y)| \leq 1$ and each $\mu_i(s_i)$ is at most p^{-1}. In more complicated cases, we end up wishing to prove that a certain integer-valued random variable with mean n^{-c} has a probability n^{-A} of exceeding a large constant C.

We are now ready to state our main conditional results. Note that in all of these it is necessary to assume that the probability q with which we choose our random set U is smaller than some positive constant δ. For colouring theorems this is not a problem, because these properties are always monotone. It is therefore enough to know that the property holds almost surely for a particular probability q to know that it holds almost surely for all probabilities larger than q.

For density theorems, we can also overcome this difficulty by partitioning any random set with large probability into a number of smaller random sets each chosen with probability less than δ. With high probability, each of these smaller random sets will satisfy the required density theorem. If we take a subset of the original set above a certain density, then this subset must have comparable density within at least one of the sets of the partition. Applying the required density theorem within this set,

we can find the required substructure, be it a k-term arithmetic progression or a complete graph of order t.

Alternatively, if we know a (robust) sparse density theorem for a small value of p, we can deduce it for a larger value q as follows. We can pick a random set $V = X_p$ by first choosing $U = X_q$ and then choosing $V = U_{p/q}$. Since the result is true for almost every $V = X_p$, it will be the case that for almost every $U = X_q$, almost every $V = U_{p/q}$ will satisfy the result. It follows by a simple averaging argument that for almost every $U = X_q$ the robust version of the density theorem holds again.

Unfortunately, for structural results, no simple argument of this variety seems to work and we will have to deal with each case as it comes.

Theorem 9.1. *Suppose that S, X and p_0 satisfy Conditions 1 and 2. Suppose also that there exist positive constants ρ and β such that for every subset $B \subset X$ of density at least ρ there are at least $\beta|S|$ sequences $s = (s_1, \ldots, s_k) \in S$ such that $s_i \in B$ for every i. Then, for any $\epsilon > 0$, there exist positive constants C and δ with the following property. Let U be a random subset of X, with elements chosen independently with probability $Cp_0 \leq q \leq \delta$. Then, with probability $1 - o(1)$, every subset A of U of density at least $\rho + \epsilon$ contains at least $(\beta - \epsilon)p^k|S|$ sequences such that $s_i \in A$ for every i.*

Proof. Basically the result is true because Theorem 4.5 proves the conclusion conditional on the four key properties set out before the statement of Theorem 4.5, and our probabilistic arguments in the last few sections show that these properties follow from Conditions 1 and 2. Indeed, we would already be done if it were not for one small extra detail: we need to deal with the fact that Theorem 4.5 has a conclusion that concerns m random sets $U_1, \ldots, U_m$, whereas we want a conclusion that concerns a single random set U.

Let η, λ, d and m be as required by Theorem 4.5. Condition 1 implies that property P2 holds with probability $1 - o(|X|^{-k})$. Lemma 8.4 tells us that property P1 holds with probability $1 - o(|X|^{-k})$ provided that Conditions 1 and 2 hold, for some α that depends on η. Property P0 plainly holds with high probability. Finally, Lemma 6.14 tells us that if properties P0, P1 and P2 hold with probability $1 - o(|X|^{-k})$, then property P3 holds with probability $1 - o(1)$. Thus, with probability $1 - o(1)$, we have all four properties.

It follows from Theorem 4.5 that if $U_1, \ldots, U_m$ are independent random subsets of X, each chosen binomially with probability $p \geq p_0$, and $\mu_1, \ldots, \mu_m$ are their associated measures, then, with probability $1 - o(1)$, $\mathbb{E}_{s \in S} h(s_1) \ldots h(s_k) \geq \beta - \epsilon$ whenever $0 \leq h \leq m^{-1}(\mu_1 + \cdots + \mu_m)$ and $\|h\|_1 \geq \rho + 3\epsilon/4$.

Let $U = U_1 \cup \cdots \cup U_m$. Then U is a random set with each element chosen independently with probability $q = 1 - (1-p)^m \geq mp(1 - \epsilon/8)$, provided δ (and hence p) is sufficiently small. Let μ be the associated measure of U, let $0 \leq f \leq \mu$ and suppose that $\|f\|_1 \geq \rho + \epsilon$. Then replacing f by the smaller function $h = \min\{f, m^{-1}(\mu_1 + \cdots + \mu_m)\}$ we have $\|h\|_1 \geq \rho + 3\epsilon/4$, which implies that $\mathbb{E}_{s \in S} h(s_1) \ldots h(s_k) \geq \beta - \epsilon$, which in turn implies that $\mathbb{E}_{s \in S} f(s_1) \ldots f(s_k) \geq \beta - \epsilon$. $\qquad \square$

Conditions 1 and 2 also imply an abstract colouring result and an abstract structural result in a very similar way.

Theorem 9.2. *Suppose that S, X and p_0 satisfy Conditions 1 and 2. Suppose also that r is a positive integer and β a positive constant such that for every colouring of X with r colours there are at least $\beta|S|$ sequences $s = (s_1, \ldots, s_k) \in S$ such that each s_i has the same colour. Then there exist positive constants C and δ with the following property. Let U be a random subset of X, with elements chosen independently with probability $Cp_0 \leq q \leq \delta$. Then, with probability $1 - o(1)$, every colouring of U with*

r colours contains at least $2^{-(k+2)}\beta p^k |S|$ sequences $s = (s_1, \ldots, s_k) \in S$ such that each s_i has the same colour and each s_i is an element of U.

The only further ingredient needed to prove this theorem is Theorem 4.8. Other than this, the proof is much the same as that of Theorem 9.1.

Theorem 9.3. *Suppose that S, X and p_0 satisfy Conditions 1 and 2 and let $\mathcal{V}$ be a collection of $2^{o(p|X|)}$ subsets of X. Then, for any $\epsilon > 0$, there exist positive constants C and δ with the following property. Let U be a random subset of X, with elements chosen independently with probability $Cp_0 \leq q \leq \delta$, and associated measure μ. Then, with probability $1 - o(1)$, for every function $f : X \to \mathbb{R}$ with $0 \leq f \leq \mu$ there exists a function $g : X \to \mathbb{R}$ with $0 \leq g \leq 1$ such that*

$$\mathbb{E}_{s \in S} f(s_1) \ldots f(s_k) \geq \mathbb{E}_{s \in S} g(s_1) \ldots g(s_k) - \epsilon$$

and, for all $V \in \mathcal{V}$,

$$|\sum_{x \in V} f(x) - \sum_{x \in V} g(x)| \leq \epsilon |X|.$$

The main extra point to note here is that Conditions 1 and 2 imply not just property P3 but also property P3′. This allows us to apply Theorem 4.10.

9.2   The critical exponent

The aim of this paper has been to prove results that are, in terms of p, best possible to within a constant. A preliminary task is to work out the probability below which we cannot hope to prove a result. (For density problems, it is easy to prove that below this probability the result is not even true. For natural colouring problems, it usually seems to be the case that the result is not true, but the known proofs are far from trivial.) To within a constant, the probability in question is the probability p such that the following holds: for each $j \leq k$ and each $x \in X$ the expected number of elements $s \in S$ such that $s_j = x$ (that is, such that $s \in S_j(x)$), and s_i belongs to X_p for each $i \neq j$ is equal to 1.

In concrete situations, X will be one of a family of sets of increasing size, and S will be one of a corresponding family of sets of sequences. Then it is usually the case that the probability p calculated above is within a constant of $|X|^{-\alpha}$ for some rational number α that does not depend on which member of the family one is talking about. In this situation, we shall call α the *critical exponent* for the family of problems. Our results will then be valid for all p that exceed $C|X|^{-\alpha}$ for some constant C. We shall denote the critical exponent by α_S, even though strictly speaking it depends not on an individual S but on the entire family of sets of sequences.

To give an example, if S consists of all non-degenerate edge-labelled copies of K_4 in K_n, then the expected number of copies with a particular edge in a particular place, given that that edge belongs to U, is $2(n-2)(n-3)p^5$ (since each $S_j(e)$ has size $2(n-2)(n-3)$ and there are five edges that must be chosen). Setting that equal to 1 tells us that p is within a constant of $n^{-2/5}$, so the critical exponent is 2/5. (This is a special case of the formula $1/m_k(K) = (v_K - k)/(e_K - 1)$.)

This calculation is exactly what we do in general: if each element of S is a sequence of length k and we are given that $x \in X_p$, then the expected number of elements of $S_j(x)$ that have all their terms in X_p is $p^{k-1}|S_j(x)|$. This equals 1 when $p = |S_j(x)|^{-1/(k-1)}$. If $|S_j(x)| = C|X|^\theta$ for some θ that is independent of the size of the problem, then the critical exponent is therefore $\theta/(k-1)$.

If we can prove a robust density theorem for S, and can show that Conditions 1 and 2 hold when $p_0 = C|X|^{-\alpha_S}$ for some constant C, then we have proved a result that is best possible to within a constant. For colouring theorems, we cannot be quite so sure that the result is best possible, but in almost all examples where the 0-statement has been proved, it does indeed give a bound of the form $c|X|^{-\alpha_S}$.

9.3 Unconditional results

In this section we concentrate on the two kinds of sequence system for which we have proved that Conditions 1 and 2 hold when $p_0 = C|X|^{-\alpha_S}$.

Recall that S *has two degrees of freedom* if $S_i(x) \cap S_j(y)$ is either empty or a singleton whenever $i \neq j$. A good example of such a system is the set of k-term arithmetic progressions in $\mathbb{Z}_p$ for some prime p. We say that S *is a set of copies of a hypergraph* if K is a k-uniform hypergraph with edges $a_1, \ldots, a_e$, X is the complete k-uniform hypergraph $K_n^{(k)}$, and S is the set of all sequences of the form $(\phi(a_1), \ldots, \phi(a_e))$, where ϕ is an injection from the vertices of K to $\{1, 2, \ldots, n\}$.

As above, we assume that S has the additional homogeneity property that $S_1(x) \cap S_k(y)$ always has the same size when it is non-empty. (In the hypergraph case, e plays the role of k and k has a different meaning: thus, the property in that case is that $S_1(x) \cap S_e(y)$ always has the same size when it is non-empty.) And in the hypergraph case, we make the further assumption that the hypergraph K is *strictly balanced*, which means that for every proper subhypergraph $J \subset K$ we have the inequality $\frac{e_J - 1}{v_J - k} < \frac{e_K - 1}{v_K - k}$. When this happens, we write $m_k(K)$ as shorthand for $\frac{e_K - 1}{v_K - k}$.

Given a system S with two degrees of freedom, let t be the size of each $S_j(x)$, and suppose that $t = |X|^\gamma$. Then the critical exponent of S is $\gamma/(k-1)$. (Note that $|X|^{-\alpha_S} = t^{-1/(k-1)}$.) When S is a set of copies of a strictly balanced hypergraph K, the critical exponent is $1/m_k(K)$. It is straightforward to show that sparse density results cannot hold for random subsets of X chosen with probability $c|X|^{-\alpha_S}$ if c is a sufficiently small positive constant. Broadly speaking, we shall show that they *do* hold for random subsets chosen with probability $C|X|^{-\alpha_S}$ when C is a sufficiently *large* positive constant.

Let us call a system S *good* if the above properties hold. That is, roughly speaking, a good system is a system with certain homogeneity properties that either has two degrees of freedom or comes from copies of a graph or hypergraph. We shall also assume that $|X|$ is sufficiently large. When we say "there exists a constant C," this should be understood to depend only on k in the case of systems of two degrees of freedom, and only on K in the case of copies of a strictly balanced hypergraph, together with parameters such as density or the number of colours in a colouring that have been previously mentioned in the statement.

Theorem 9.4. *Let X be a finite set and let S be a good system of ordered subsets of X. Suppose that there exist positive constants ρ and β such that for every subset $B \subset X$ of density at least ρ there are at least $\beta|S|$ sequences $s = (s_1, \ldots, s_k) \in S$ such that $s_i \in B$ for every i. Then, for any $\epsilon > 0$, there exist positive constants C and δ with the following property. Let U be a random subset of X, with elements chosen independently with probability $C|X|^{-\alpha_S} \leq p \leq \delta$. Then, with probability $1 - o(1)$, every subset A of U of order at least $(\rho + \epsilon)|U|$ contains at least $(\beta - \epsilon)p^k|S|$ sequences such that $s_i \in A$ for every i.*

Proof. By Theorem 9.1, all we have to do is check Conditions 1 and 2. Condition 1 is given to us by Lemma 7.2 when S has two degrees of freedom, and by Lemma 7.4 when S is a system of copies

of a graph or hypergraph, even when $C = 1$. (In the case where S has two degrees of freedom, see the remarks following Lemma 7.2 for an explanation of why the result implies Condition 1 when $p = |X|^{-\alpha_S}$.)

When S has two degrees of freedom, Condition 2 holds as long as $p^{-(k-2)} \leq \alpha pt$, as we have already remarked. This tells us that p needs to be at least $(\alpha t)^{-1/(k-1)}$. In this case, $t = |S_j(x)|$ for each x and j, so $(\alpha t)^{-1/(k-1)}$ is within a constant of $|X|^{-\alpha_s}$, as required. When S comes from copies of a strictly balanced graph or hypergraph, Lemmas 8.5 and 8.7 give us Condition 2, again with $p = C|X|^{-\alpha_S}$. $\quad\square$

Exactly the same proof (except that we use Theorem 9.2 instead of Theorem 9.1) gives us the following general sparse colouring theorem.

Theorem 9.5. *Let X be a finite set and let S be a good system of ordered subsets of X. Suppose that r is a positive integer and that β is a positive constant such that for every colouring of X with r colours there are at least $\beta|S|$ sequences $s = (s_1, \ldots, s_k) \in S$ such that each s_i has the same colour. Then there exist positive constants C and δ with the following property. Let U be a random subset of X, with elements chosen independently with probability $C|X|^{-\alpha_S} \leq p \leq \delta$. Then, with probability $1 - o(1)$, every colouring of U with r colours contains at least $2^{-(k+2)}\beta p^k|S|$ sequences $s = (s_1, \ldots, s_k) \in S$ such that each s_i has the same colour and each s_i is an element of U.*

Finally, we have the following general sparse structural theorem.

Theorem 9.6. *Let X be a finite set and let S be a good system of ordered subsets of X. Then, for any $\epsilon > 0$, there exist positive constants C and δ with the following property. Let U be a random subset of X, with elements chosen independently with probability $C|X|^{-\alpha_S} \leq p \leq \delta$, let μ be the associated measure of U and let $\mathcal{V}$ be a collection of $2^{o(p|X|)}$ subsets of X. Then, with probability $1 - o(1)$, for every function f with $0 \leq f \leq \mu$ there exists a function g with $0 \leq g \leq 1$ such that*

$$\mathbb{E}_{s \in S} f(s_1) \ldots f(s_k) \geq \mathbb{E}_{s \in S} g(s_1) \ldots g(s_k) - \epsilon$$

and, for all $V \in \mathcal{V}$,

$$|\sum_{x \in V} f(x) - \sum_{x \in V} g(x)| \leq \epsilon|X|.$$

In applications, we often want g to take values in $\{0, 1\}$ rather than $[0, 1]$. This can be achieved by a simple and standard modification of the above result.

Corollary 9.7. *Let X be a finite set and let S be a good system of ordered subsets of X. Then, for any $\epsilon > 0$, there exist positive constants C and δ with the following property. Let U be a random subset of X, with elements chosen independently with probability $C|X|^{-\alpha_S} \leq p \leq \delta$, let μ be the associated measure of U and let $\mathcal{V}$ be a collection of $2^{o(p|X|)}$ subsets of X. Then, with probability $1 - o(1)$, for every function f with $0 \leq f \leq \mu$ there exists a function h taking values in $\{0, 1\}$ such that*

$$\mathbb{E}_{s \in S} f(s_1) \ldots f(s_k) \geq \mathbb{E}_{s \in S} h(s_1) \ldots h(s_k) - \epsilon$$

and, for all $V \in \mathcal{V}$,

$$|\sum_{x \in V} f(x) - \sum_{x \in V} h(x)| \leq \epsilon|X|.$$

Proof. The basic idea of the argument is to choose a function g that satisfies the conclusion of Theorem 9.6 with ϵ replaced by $\epsilon/2$, and to let $h(x) = 1$ with probability $g(x)$ and 0 with probability $1 - g(x)$, all choices being made independently. Then concentration of measure tells us that with high probability the estimates are not affected very much.

Note first that the expectation of $\mathbb{E}_{s \in S} h(s_1) \ldots h(s_k)$ is $\mathbb{E}_{s \in S} g(s_1) \ldots g(s_k)$. By how much can changing the value of $h(x)$ change the value of $\mathbb{E}_{s \in S} h(s_1) \ldots h(s_k)$? Well, if x is one of $s_1, \ldots, s_k$ then $h(s_1) \ldots h(s_k)$ can change by at most 1 and otherwise it does not change. The probability that x is one of $s_1, \ldots, s_k$ is $k/|X|$, by the homogeneity of S (which tells us that each s_j is uniformly distributed). By Azuma's inequality, it follows that the probability that $|\mathbb{E}_{s \in S} g(s_1) \ldots g(s_k) - \mathbb{E}_{s \in S} h(s_1) \ldots h(s_k)| \geq \epsilon/2$ is at most $2 \exp(-\epsilon^2 |X|/8k^2)$. This gives us the first conclusion with very high probability.

The second is obtained in a similar way. For each $V \in \mathcal{V}$ the probability that $|\sum_{x \in V} h(x) - \sum_{x \in V} g(x)| \geq \epsilon|X|/2$ is, by Azuma's inequality, at most $2 \exp(-\epsilon^2 |X|/8)$. Since there are $2^{o(p|X|)}$ sets in $\mathcal{V}$, a union bound gives the second conclusion with very high probability as well. $\square$

10 Applications

10.1 Density results

As a first example, let us prove Theorem 1.12, the sparse analogue of Szemerédi's theorem. We shall consider Szemerédi's theorem as a statement about arithmetic progressions mod p in dense subsets of $\mathbb{Z}_p$ for a prime p. We do this because the set of k-term arithmetic progressions in $\mathbb{Z}_p$ is a homogeneous system with two degrees of freedom. However, once we have the result for this version of Szemerédi's theorem, we can easily deduce it for the more conventional version concerning a sparse random subset of $[n]$. We simply choose a prime p between $2n$ and $4n$, pick a sparse random subset U of $\mathbb{Z}_p$, and then apply the result to subsets of U that happen to be subsets of $\{1, 2, \ldots, n\}$, since arithmetic progressions in these subsets will not wrap around. Similar arguments allow us to replace $[n]$ by $\mathbb{Z}_p$ for our later applications, so we mention once and for all now that for each application it is easy to deduce from the result we state a result for sparse subsets of intervals (or grids in the multidimensional case).

Since we wish to use the letter p to denote a probability, we shall now let n be a large prime.

By Theorem 9.1, all we have to do is check the robust version of Szemerédi's theorem, which can be proved by a simple averaging argument, originally observed by Varnavides [64] (who stated it for 3-term progressions).

Theorem 10.1. *Let k be an integer and $\delta > 0$ a real number. Then there exists an integer n_0 and $c > 0$ such that if n is a prime greater than or equal to n_0 and B is a subset of $\mathbb{Z}_n$ with $|B| \geq \delta n$ then B contains at least cn^2 k-term arithmetic progressions.*

Proof. Let m be such that every subset of $\{1, 2, \ldots, m\}$ of density $\delta/2$ contains a k-term arithmetic progression. Now let B be a subset of $\mathbb{Z}_n$ of density δ. For each a and d with $d \neq 0$, let $P_{a,d}$ be the mod-n arithmetic progression $\{a, a + d, \ldots, a + (m-1)d\}$. If we choose $P_{a,d}$ at random, then the expected density of B inside $P_{a,d}$ is δ, so with probability at least $\delta/2$ it is at least $\delta/2$. It follows with probability at least $\delta/2$ that $P_{a,d}$ contains an arithmetic progression that is contained in B. Since $P_{a,d}$ contains at most $m(m-1)$ k-term arithmetic progressions, it follows that with probability at least $\delta/2$ at least $1/m(m-1)$ of the k-term progressions inside $P_{a,d}$ are contained in B. But every k-term

arithmetic progression is contained in the same number of progressions $P_{a,d}$. Therefore, the number of progressions in B is at least $\frac{\delta}{2} \frac{n(n-1)}{m(m-1)} \geq \frac{\delta}{2} \frac{n^2}{m^2}$. □

Very similar averaging arguments are used to prove the other robust density results we shall need in this subsection, so we shall be sketchy about the proofs and sometimes omit them altogether.

The next result is the sparse version of Szemerédi's theorem. Recall that we write X_p for a random subset of X where each element is chosen independently with probability p, and we say that a set I is (δ, k)-Szemerédi if every subset of I with cardinality at least $\delta|I|$ contains a k-term arithmetic progression.

Theorem 10.2. *Given $\delta > 0$ and a natural number $k \geq 3$, there exists a constant C such that if $p \geq Cn^{-1/(k-1)}$, then the probability that $(\mathbb{Z}_n)_p$ is (δ, k)-Szemerédi is $1 - o(1)$.*

Proof. In the case where p is not too large, this follows immediately from Theorems 9.4 and 10.1. The result for larger probabilities can be deduced by using the argument given before Theorem 9.1. Alternatively, note that a subset of relative density δ within a subset of $[n]_p$ has density δp in $[n]$. So if p is larger than a fixed constant λ (as it will be in the case not already covered by Theorem 9.4), we can just apply Szemerédi's theorem itself. □

A simple corollary of Theorem 10.2 is a sparse analogue of van der Waerden's theorem [63] on arithmetic progressions in r-colourings of $[n]$. Note that this theorem was proved much earlier by Rödl and Ruciński [47] and is known to be tight.

Let us now prove sparse versions of two generalizations of Szemerédi's theorem. The first generalization is the multidimensional Szemerédi theorem, due to Furstenberg and Katznelson [15]. We shall state it in its robust form, which is in fact the statement that Furstenberg and Katznelson directly prove. (It also follows from the non-robust version by means of an averaging argument.)

Theorem 10.3. *Let r be a positive integer and $\delta > 0$ a real number. If $P \subset \mathbb{Z}^r$ is a fixed set, then there is a positive integer n_0 and a constant $c > 0$ such that, for $n \geq n_0$, every subset B of the grid $[n]^r$ with $|B| \geq \delta n^r$ contains cn^{r+1} subsets of the form $a + dP$, where $a \in [n]^r$ and d is a positive integer.*

Just as with Szemerédi's theorem, this statement is equivalent to the same statement for subsets of $\mathbb{Z}_n^r$. So let P be a subset of $\mathbb{Z}^r$ and let $(x_1, \ldots, x_k)$ be an ordering of the elements of P. Let S be the set of sequences of the form $s_{a,d} = (a + x_1 d, \ldots, a + x_k d)$ with $d \neq 0$. Then S is homogeneous and has two degrees of freedom. Moreover, if n is large enough, then no two elements of $s_{a,d}$ are the same. From the conclusion of Theorem 10.3 it follows that there are at least $c|S|$ sequences in S with all their terms in B. We have therefore checked all the conditions for Theorem 9.4, so we have the following sparse version of the multidimensional Szemerédi theorem. (As before, the result for larger p follows easily from the result for smaller p.) We define a subset I of $\mathbb{Z}_n^r$ to be (δ, P)-Szemerédi if every subset of I with cardinality at least $\delta|I|$ contains a homothetic copy $a + dP$ of P.

Theorem 10.4. *Given integers r and k, a real number $\delta > 0$ and a subset $P \subset \mathbb{Z}^r$ of order k, there exists a constant C such that if $p \geq Cn^{-1/(k-1)}$, then the probability that $(\mathbb{Z}_n^r)_p$ is (δ, P)-Szemerédi is $1 - o(1)$.*

The second generalization of Szemerédi's theorem we wish to look at is the polynomial Szemerédi theorem of Bergelson and Leibman [2]. Their result is the following.

Theorem 10.5. *Let $\delta > 0$ be a real number, let k be a positive integer and let $P_1, \ldots, P_k$ be polynomials with integer coefficients that vanish at zero. Then there exists an integer n_0 such that if $n \geq n_0$ and B is a subset of $[n]$ with $|B| \geq \delta n$ then B has a subset of the form $\{a, a + P_1(d), \ldots, a + P_k(d)\}$.*

We will focus on the specific case where the polynomials are $x^r, 2x^r, \ldots, (k-1)x^r$ (so k has been replaced by $k-1$). In this case, the theorem tells us that we can find a k-term arithmetic progression with common difference that is a perfect rth power. We restrict to this case, because it is much easier to state and prove an appropriate robust version for this case than it is for the general case.

Note that if $a, a + d^r, \cdots, a + (k-1)d^r \in [n]$, then $d \leq (n/k)^{1/r}$. This observation and another easy averaging argument enable us to replace Theorem 10.5 by the following equivalent robust statement about subsets of $\mathbb{Z}_n$ (see, for example, [25]).

Corollary 10.6. *Let k, r be integers and $\delta > 0$ a real number. Then there exists an integer n_0 and a constant $c > 0$ such that if $n \geq n_0$ and B is a subset of $\mathbb{Z}_n$ with $|B| \geq \delta n$ then B contains at least $cn^{1+1/r}$ pairs (a, d) such that $a, a + d^r, \cdots, a + (k-1)d^r \in B$ and $d \leq (n/k)^{1/r}$.*

Let us say that a subset I of $\mathbb{Z}_n$ is (δ, k, r)-*Szemerédi* if every subset of I with cardinality at least $\delta|I|$ contains a k-term progression of the form $a, a + d^r, \ldots, a + (k-1)d^r$ with $d \leq (n/k)^{1/r}$.

Theorem 10.7. *Let k, r be integers and $\delta > 0$ a real number. Then there exists a constant C such that if $p \geq Cn^{-1/(k-1)r}$, then the probability that $(\mathbb{Z}_n)_p$ is (δ, k, r)-Szemerédi is $1 - o(1)$.*

Proof. Let $X = \mathbb{Z}_n$ and S be the collection of progressions of the form $a, a + d^r, \ldots, a + (k-1)d^r$ with $d \leq (n/k)^{1/r}$. Because of this restriction on d, S has two degrees of freedom. It is also obviously homogeneous. The size of each $S_j(x)$ is $n^{1/r}$ to within a constant, so the critical exponent is $\gamma/(k-1)$ with $\gamma = 1/r$. Therefore, provided p is at most some constant λ, the result follows from Theorem 9.4 and Corollary 10.6. For p larger than λ, the result follows from the polynomial Szemerédi theorem itself. $\qquad\square$

Note that the particular case of this theorem when $k = 2$ was already proved by Nguyen [41]. To see that this result is sharp, note that the number of k-term progressions with rth power difference in the random set is roughly $p^k n^{1+1/r}$. This is smaller than the number of vertices pn when $p = n^{-1/(k-1)r}$.

We will now move on to proving sparse versions of Turán's theorem for strictly k-balanced k-uniform hypergraphs. As we mentioned in the introduction, some of the dense results are not known, but this does not matter to us, since our aim is simply to show that whatever results can be proved in the dense case carry over to the sparse random case when the probability exceeds the critical probability.

For a k-uniform hypergraph K, let $\mathrm{ex}(n, K)$ denote the largest number of edges a subgraph of $K_n^{(k)}$ can have without containing a copy of K. As usual, we need a robust result that says that once a graph has more edges than the extremal number for K, by a constant proportion of the total number of edges in $K_n^{(k)}$, then it must contain many copies of K. The earliest version of such a supersaturation result was proved by Erdős and Simonovits [8]. The proof is another easy averaging argument along the lines of the proof of Theorem 10.1.

Theorem 10.8. *Let K be a k-uniform hypergraph. Then, for any $\epsilon > 0$, there exists $\delta > 0$ such that if L is a k-uniform hypergraph on n vertices and*

$$e(L) > \mathrm{ex}(n, K) + \epsilon n^k,$$

then L contains at least δn^{v_K} copies of K.

Let $\pi_k(K)$ be the limit as n tends to infinity of $\mathrm{ex}(n,K)/\binom{n}{k}$. We will say that a k-uniform hypergraph H is (K,ϵ)-Turán if any subset of the edges of H of size

$$(\pi_k(K) + \epsilon)e(H)$$

contains a copy of K. Recall that $G_{n,p}^{(k)}$ is a random k-uniform hypergraph on n vertices, where each edge is chosen with probability p, and when K is strictly k-balanced $m_k(K) = (e_K - 1)/(v_K - k)$.

Theorem 10.9. *For every $\epsilon > 0$ and every strictly k-balanced k-uniform hypergraph K, there exists a constant C such that if $p \geq Cn^{-1/m_k(K)}$, then the probability that $G_{n,p}^{(k)}$ is (K,ϵ)-Turán is $1 - o(1)$.*

Proof. For p smaller than a fixed constant λ, the result follows immediately from Theorems 9.4 and 10.8. For $p \geq \lambda$, we may apply the argument discussed before Theorem 9.1. That is, we may partition $G_{n,p}^{(k)}$ into a small set of random graphs, each of which has density less than λ and each of which is (K,ϵ)-Turán. If now we have a subgraph of $G_{n,p}^{(k)}$ of density at least $\frac{\mathrm{ex}(n,K)}{\binom{n}{k}} + \epsilon$, then this subgraph must have at least this density in one of the graphs from the partition. Applying the fact that this subgraph is (K,ϵ)-Turán implies the result. □

In particular, this implies Theorem 1.10, which is the particular case of this theorem where K is a strictly balanced graph. Then $\mathrm{ex}(n,K)$ is known to be $\left(1 - \frac{1}{\chi(K)-1} + o(1)\right)\binom{n}{2}$, where $\chi(K)$ is the chromatic number of K.

10.2   Colouring results

We shall now move on to colouring results that do not follow from their corresponding density versions. Let us begin with Ramsey's theorem. As ever, the main thing we need to check is that a suitable robust version of the theorem holds. And indeed it does: it is a very simple consequence of Ramsey's theorem that was noted by Erdős [7].

Theorem 10.10. *Let H be a hypergraph and let r be a positive integer. Then there exists an integer n_0 and a constant $c > 0$ such that, if $n \geq n_0$, any colouring of the edges of $K_n^{(k)}$ with r colours is guaranteed to contain cn^{v_H} monochromatic copies of H.*

Once again the proof is the obvious averaging argument: choose m such that if the edges of $K_m^{(k)}$ are coloured with r colours, there must be a monochromatic copy of H, and then a double count shows that for every r-colouring of the edges of $K_n^{(k)}$ there are at least $\binom{n}{v_H}/\binom{m}{v_H}$ monochromatic copies of H.

Recall that, given a k-uniform hypergraph K and a natural number r, a hypergraph is (K,r)-*Ramsey* if every r-colouring of its edges contains a monochromatic copy of K. We are now ready to prove Theorem 1.9, which for convenience we restate here.

Theorem 10.11. *Given a natural number r and a strictly k-balanced k-uniform hypergraph K, there exists a positive constant C such that if $p \geq Cn^{-1/m_k(K)}$, then the probability that $G_{n,p}^{(k)}$ is (K,r)-Ramsey is $1 - o(1)$.*

Proof. For a sufficiently large constant C, the result for $p = Cn^{-1/m_k(K)}$ follows from Theorems 9.5 and 10.10. For $q > p$, the result follows from the monotonicity of the Ramsey property. To see this, choose a random hypergraph $G_{n,q}^{(k)}$ and then choose a subhypergraph by randomly selecting each edge of $G_{n,q}^{(k)}$ with probability p/q. The resulting hypergraph is distributed as $G_{n,p}^{(k)}$, so with probability $1 - o(1)$ it is (K, r)-Ramsey. But then any r-colouring of $G_{n,q}^{(k)}$ will yield an r-colouring of this $G_{n,p}^{(k)}$, which always contains a monochromatic copy of K. $\square$

With only slightly more effort we can obtain a robust conclusion. Theorem 9.5 tells us that with high probability the number of monochromatic copies of K in any r-colouring of $G_{n,p}^{(k)}$ is $cp^{e_K} n^{v_K}$ for some constant $c > 0$, and then an averaging argument implies that with high probability the number of monochromatic copies in an r-colouring of $G_{n,q}^{(k)}$ is $cq^{e_K} n^{v_K}$.

We shall now take a look at Schur's theorem [56], which states that if the set $\{1, \ldots, n\}$ is r-coloured, then there exist monochromatic subsets of the form $\{x, y, x+y\}$. As with our results concerning Szemerédi's theorem, it is more convenient to work in $\mathbb{Z}_n$. To see that this implies the equivalent theorem in $[n]$, let $[n]_p$ be a random subset of $[n]$ made from the union of two smaller random subsets, each chosen with probability q such that $p = 2q - q^2$. Call these sets U_1 and U_2. Then the subset of $\mathbb{Z}_{2n}$ formed by placing the set U_1 in the position $\{1, \ldots, n\}$ and the set $-U_2$ in the set $\{-1, \ldots, -n\}$ (the overlap $n = -n$ is irrelevant to the argument, since it is unlikely to be in the set) will produce a random subset of $\mathbb{Z}_{2n}$ where each element is chosen with probability q. If a sparse version of Schur's theorem holds in $\mathbb{Z}_{2n}$, then with high probability, any $2r$-colouring of this random set yields $cq^3 n^2$ monochromatic sets $\{x, y, x+y\}$ for some constant $c > 0$.

Consider now an r-colouring of the original set $U_1 \cup U_2$ in r colours $C_1, \ldots, C_r$. This induces a colouring of $U_1 \cup -U_2 \subset \mathbb{Z}_{2n}$ with $2r$ colours $C_1, \ldots, C_{2r}$: if $x \in U_1$ and is coloured with colour C_i in $[n]$, then we continue to colour it with colour C_i, whereas if $x \in -U_2$ and $-x$ has colour C_i in $[n]$, then we colour it with colour C_{i+r}. We have already noted that this colouring must contain many monochromatic sets $\{x, y, x+y\}$, and each one corresponds to a monochromatic set (either $\{x, y, x+y\}$ or $\{-x, -y, -(x+y)\}$) in the original colouring.

The robust version of Schur's theorem can be deduced from one of the standard proofs, which itself relies on Ramsey's theorem for triangles and many colours.

Theorem 10.12. *Let r be a positive integer. Then there exists an integer n_0 and a constant c such that, if $n \geq n_0$, any r-colouring of $\{1, \ldots, n\}$ contains at least cn^2 monochromatic triples of the form $\{x, y, x+y\}$.*

We shall say that a subset I of the integers is r-*Schur* if for every r-colouring of the points of I there is a monochromatic triple of the form $\{x, y, x+y\}$. The $r = 2$ case of the following theorem was already known: it is a result of Graham, Rödl and Ruciński [23].

Theorem 10.13. *For every positive integer r there exists a constant C such that if $p \geq Cn^{-1/2}$, then the probability that $(\mathbb{Z}_n)_p$ is r-Schur is $1 - o(1)$.*

Proof. Let $X = \mathbb{Z}_n \setminus \{0\}$ and S be the collection of subsets of X of the form $\{x, y, x+y\}$ with all of x, y and $x+y$ distinct. Since any two of x, y and $x+y$ determine the third, it follows that $|S_i(a) \cap S_j(b)| \leq 1$ whenever $i, j \in \{1, 2, 3\}$, $i \neq j$, and $a, b \in X$. Therefore, S has two degrees of freedom. Furthermore, each $S_i(a)$ has size $n - 3$. By Theorem 10.12, there exists a constant c such that, for n sufficiently large, any r-colouring of $\mathbb{Z}_n$ contains at least cn^2 monochromatic subsets of the form $\{x, y, x+y\}$.

Applying Theorem 9.5, we see that there exist positive constants C and c' such that, with probability $1 - o(1)$ a random subset U of $\mathbb{Z}_n$ chosen with probability $p = Cn^{-1/2}$ satisfies the condition that, in any r-colouring of U, there are at least $c'p^3n^2$ monochromatic subsets of the form $\{x, y, x + y\}$. In particular, U is r-Schur. Once again, the result for larger probabilities follows easily. □

As we mentioned in the introduction, it is quite a bit harder to prove 0-statements for colouring statements than it is for density statements. However, 0-statements for partition regular systems have been considered in depth by Rödl and Ruciński [48], and their result implies that Theorem 10.13 is sharp.

A far-reaching generalization of Schur's theorem was proved by Rado [42]. It is likely that our methods could be used to prove other cases of Rado's theorem, but we have not tried to do so here, since we would have to impose a condition on the configurations analogous to the strictly balanced condition for graphs and hypergraphs.

10.3 The hypergraph removal lemma

Rather than jumping straight into studying hypergraphs, we shall begin by stating a slight strengthening of the triangle removal lemma for graphs. This strengthening follows from its proof via Szemerédi's regularity lemma and gives us something like the "robust" version we need in order to use our methods to obtain a sparse result. If G is a graph and X and Y are sets of vertices, we shall write $G(X, Y)$ for the set of edges that join a vertex in X to a vertex in Y, $e(X, Y)$ for the cardinality of $G(X, Y)$ and $d(X, Y)$ for $e(X, Y)/|X||Y|$.

Theorem 10.14. *For every $a > 0$ there exists a constant K with the following property. For every graph G with n vertices, there is a partition of the vertices of G into $k \leq K$ sets $V_1, \ldots, V_k$, each of size either $\lfloor n/k \rfloor$ or $\lceil n/k \rceil$, and a set E of edges of G with the following properties.*

1. *The number of edges in E is at most an^2.*

2. *E is a union of sets of the form $G(V_i, V_j)$.*

3. *E includes all edges that join a vertex in V_i to another vertex in the same V_i.*

4. *Let G' be G with the edges in E removed. For any h, i, j, if there are edges in all of $G'(V_h, V_i)$, $G'(V_i, V_j)$ and $G'(V_h, V_j)$, then the number of triangles xyz with $x \in V_h$, $y \in V_i$ and $z \in V_j$ is at least $a^3|V_h||V_i||V_j|/128$.*

In particular, this tells us that after we remove just a few edges we obtain a graph that contains either no triangles or many triangles. Let us briefly recall the usual statement of the dense triangle removal lemma and see how it follows from Theorem 10.14.

Corollary 10.15. *For every $a > 0$ there exists a constant $c > 0$ with the following property. For every graph G with n vertices and at most cn^3 triangles it is possible to remove at most an^2 edges from G in such a way that the resulting graph contains no triangles.*

Proof. Apply Theorem 10.14 to a and let $c = a^3/200K^3$. Now let G be a graph with n vertices. Let $V_1, \ldots, V_k$ and E be as given by Theorem 10.14 and remove from G all edges in E. If we do this, then by Condition 1 we remove at most an^2 edges from G. If there were any triangle left in G, then by

Condition 4 there would have to be at least $a^3\lfloor n/k\rfloor^3/128 > cn^3$ triangles left in G, a contradiction. This implies the result. $\qquad\square$

Here now is a sketch of how to deduce a sparse triangle removal lemma from Theorem 10.14. We begin by proving a sparse version of Theorem 10.14 itself. Given a random graph U with edge probability $p \geq Cn^{-1/2}$, for sufficiently large C, let H be a subgraph of U. Now use Corollary 9.7 to find a dense graph G such that the triangle density of G is roughly the same as the relative triangle density of H in U (that is, if H has $\alpha p^3 n^3$ triangles, then G has roughly αn^3 triangles) and such that for every pair of reasonably large sets X, Y of vertices the density $d_G(X, Y)$ is roughly the same as the relative density of H inside $U(X, Y)$ (that is, the number of edges of $G(X, Y)$ is roughly p^{-1} times the number of edges of $H(X, Y)$).

Now use Theorem 10.14 to find a partition of the vertex set of G (which is also the vertex set of H) into sets $V_1, \ldots, V_k$ and to identify a set E_G of edges to remove from G. By Condition 2, E_G is a union of sets of the form $G(V_i, V_j)$. Define E_H to be the union of the corresponding sets $H(V_i, V_j)$ and remove all edges in E_H from H. If it happens that $G(V_i, V_j)$ is empty, then adopt the convention that we remove all edges from $H(V_i, V_j)$. Note that because the relative densities in dense complete bipartite graphs are roughly the same, the number of edges in E_H is at most $2apn^2$. Let G' be G with the edges in E_G removed and let H' be H with the edges in E_H removed.

Suppose now that H' contains a triangle xyz and suppose that $x \in V_h$, $y \in V_i$ and $z \in V_j$. Then none of $G'(V_h, V_i)$, $G'(V_i, V_j)$ and $G'(V_h, V_j)$ is empty, by our convention above, so Condition 4 implies that G' contains at least $a^3|V_h||V_i||V_j|/128$ triangles with $x \in V_h$, $y \in V_i$ and $z \in V_j$. Since triangle densities are roughly the same, it follows that H' contains at least $a^3p^3|V_h||V_i||V_j|/256$ triangles.

Roughly speaking, what this tells us is that Theorem 10.14 transfers to a sparse random version. From that it is easy to deduce a sparse random version of Corollary 10.15. However, instead of giving the full details of this, we shall prove (in a very similar way) a more general theorem, namely a sparse random version of the simplex removal lemma for hypergraphs, usually known just as the hypergraph removal lemma.

The dense result is due to Nagle, Rödl, Schacht and Skokan [40, 51], and independently to the second author [21]. A gentle introduction to the hypergraph removal lemma that focuses on the case of 3-uniform hypergraphs can be found in [20]. The result is as follows.

Theorem 10.16. *For every $\delta > 0$ and every integer $k \geq 2$, there exists a constant $\epsilon > 0$ such that, if G is a k-uniform hypergraph containing at most ϵn^{k+1} copies of $K_{k+1}^{(k)}$, it may be made $K_{k+1}^{(k)}$-free by removing at most δn^k edges.*

A *simplex* is a copy of $K_{k+1}^{(k)}$. As in the case of graphs, where simplices are triangles, it will be necessary to state a rather more precise and robust result. This is slightly more complicated to do than it was for graphs. However, it is much *less* complicated than it might be: it turns out not to be necessary to understand the statement of the regularity lemma for hypergraphs.

Let us make the following definition. Let H be a k-uniform hypergraph, and let $J_1, \ldots, J_k$ be disjoint $(k-1)$-uniform hypergraphs with the same vertex set as H. We shall define $H(J_1, \ldots, J_k)$ to be the set of all edges $A = \{a_1, \ldots, a_k\} \in H$ such that $\{a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_k\} \in J_i$ for every i. (Note that if $k = 2$ then the sets J_1 and J_2 are sets of vertices, so we are obtaining the sets $G(X, Y)$ defined earlier.)

Now suppose that we have a simplex in H with vertex set $(x_1, \ldots, x_{k+1})$. For every subset $\{u, v\}$ of $[k+1]$ of size 2, let us write J_{uv} for the (unique) set J_i that contains the $(k-1)$-set

$\{x_j : j \notin \{u,v\}\}$. Then for each u the set $H(J_{u1}, \ldots, J_{u,u-1}, J_{u,u+1}, \ldots, J_{u,k+1})$ is non-empty. We make this remark in order to make the statement of the next theorem slightly less mysterious. It is an analogue for k-uniform hypergraphs of Theorem 10.14. For convenience, we shall abbreviate $H(J_{u1}, \ldots, J_{u,u-1}, J_{u,u+1}, \ldots, J_{u,k+1})$ by $H(J_{uv} : v \in [k+1], v \neq u)$. (It might seem unnecessary to write "$v \in [k+1]$" every time. We do so to emphasize the asymmetry: the set depends on u, while v is a dummy variable.)

Theorem 10.17. *For every $a > 0$ there exists a constant K with the following property. For every k-uniform hypergraph H with vertex set $[n]$, there is a partition of $\binom{[n]}{k-1}$ into at most K subsets $J_1, \ldots, J_m$, with sizes differing by a factor of at most 2, and a set E of edges of H with the following properties.*

1. *The number of edges in E is at most an^k.*

2. *E is a union of sets of the form $H(J_{i_1}, \ldots, J_{i_k})$.*

3. *E includes all edges in any set $H(J_{i_1}, \ldots, J_{i_k})$ for which two of the i_h are equal.*

4. *Let H' be H with the edges in E removed. Suppose that for each pair of unequal integers $u, v \in [k+1]$ there is a set J_{uv} from the partition such that the hypergraphs $H'(J_{uv} : v \in [k+1], v \neq u)$ are all non-empty. Then the number of simplices with vertices $(x_1, \ldots, x_{k+1})$ such that the edge $(x_1, \ldots, x_{u-1}, x_{u+1}, \ldots, x_{k+1})$ belongs to $H'(J_{uv} : v \in [k+1], v \neq u)$ for every u is at least $(1/2)(a/4)^{k+1} c_K n^{k+1}$, where c_K is a constant that depends on K (and hence on a).*

Let us now convert this result into a sparse version.

Theorem 10.18. *For every $a > 0$ there exist constants C, K and δ with the following property. Let U be a random k-uniform hypergraph with vertex set $[n]$, and with each edge chosen independently with probability $Cn^{-1/k} \leq p \leq \delta$. Then with probability $1 - o(1)$ the following result holds. For every k-uniform hypergraph $F \subset U$, there is a partition of $\binom{[n]}{k-1}$ into at most K subsets $J_1, \ldots, J_m$, with sizes differing by a factor of at most 2, and a set E_F of edges of F with the following properties.*

1. *The number of edges in E_F is at most apn^k.*

2. *E_F is a union of sets of the form $F(J_{i_1}, \ldots, J_{i_k})$.*

3. *E_F includes all edges in any set $F(J_{i_1}, \ldots, J_{i_k})$ for which two of the i_h are equal.*

4. *Let F' be F with the edges in E_F removed. Suppose that for each pair of unequal integers $u, v \in [k+1]$ there is a set J_{uv} from the partition such that the hypergraphs $F'(J_{uv} : v \in [k+1], v \neq u)$ are all non-empty. Then the number of simplices with vertices $(x_1, \ldots, x_{k+1})$ such that the edge $(x_1, \ldots, x_{u-1}, x_{u+1}, \ldots, x_{k+1})$ belongs to $F'(J_{uv} : v \in [k+1], v \neq u)$ for every u is at least $(1/4)(a/8)^{k+1} c_K p^{k+1} n^{k+1}$, where c_K is a constant that depends on K.*

Proof. We have essentially seen the argument in the case of graphs. To start with, let us apply Corollary 9.7 with S as the set of labelled simplices, f as p^{-1} times the characteristic function of F, $\mathcal{V}$ as the collection of all sets of the form $K_n^{(k)}(J_1, \ldots, J_k)$ where each J_i is a collection of sets of size $k - 1$ (that is, the set of ordered sequences of length k in $[n]$ such that removing the ith vertex gives

61

you an element of J_i), and $\epsilon = (1/4)(a/8)^{k+1}c_K$, where K and c_K come from applying Theorem 10.17 with $a/2$ rather than a. With this choice, ϵ will also be less than $a/2K^k$.

Note that $\alpha_S = 1/k$ in this case, and that the cardinality of $\mathcal{V}$ is at most $2^{kn^{k-1}}$, so the corollary applies. From that we obtain a hypergraph H (with characteristic function equal to the function h provided by the corollary) such that $p^{-(k+1)}$ times the number of simplices in F is at least the number of simplices in H minus ϵn^{k+1}, and such that the number of edges in $H(J_1, \ldots, J_k)$ differs from p^{-1} times the number of edges in $F(J_1, \ldots, J_k)$ by at most ϵn^k for every $(J_1, \ldots, J_k)$.

We now apply Theorem 10.17 to H with a replaced by $a/2$. Let E_H be the set of edges that we obtain and let H' be H with the edges in E_H removed.

Let $J_1, \ldots, J_m$ be the sets that partition $\binom{[n]}{k-1}$, and remove all edges from F that belong to a set $F(J_{i_1}, \ldots, J_{i_k})$ such that the edges of $H(J_{i_1}, \ldots, J_{i_k})$ belong to E_H (including when $H(J_{i_1}, \ldots, J_{i_k})$ is empty). Let E_F be the set of removed edges and let F' be F after the edges are removed.

Since $m \le K$, there are at most K^k k-tuples $(J_{i_1}, \ldots, J_{i_k})$. For each such k-tuple the number of edges in $H(J_{i_1}, \ldots, J_{i_k})$ differs from p^{-1} times the number of edges in $F(J_{i_1}, \ldots, J_{i_k})$ by at most ϵn^k. Therefore, since E_H and E_F are unions of sets of the form $H(J_{i_1}, \ldots, J_{i_k})$ and $F(J_{i_1}, \ldots, J_{i_k})$, respectively, and since $|E_H| \le an^k/2$, it follows that $|E_F| \le (a/2 + \epsilon K^k)pn^k \le apn^k$. This gives us Condition 1. Conditions 2 and 3 are trivial from the way we constructed E_F. So it remains to prove Condition 4.

Suppose, then, that for all $u, v \in [k+1]$ there is a set J_{uv} such that there are edges in all of the hypergraphs $F'(J_{uv} : v \in [k+1], v \ne u)$ for $u = 1, 2, \ldots, k+1$. Then there must be edges in all the hypergraphs $H'(J_{uv} : v \in [k+1], v \ne u)$ as well, or we would have removed the corresponding sets of edges from F. By Condition 4 of the dense result applied to H, it follows that H' contains at least $(1/2)(a/8)^{k+1}c_K n^{k+1}$ simplices, which implies that H does as well, which implies that F contains at least $((1/2)(a/8)^{k+1}c_K - \epsilon)p^{k+1}n^{k+1}$ simplices, which gives us the bound stated. $\qquad\square$

Now let us deduce the simplex removal lemma. This is just as straightforward as it was for graphs.

Corollary 10.19. *For every $a > 0$ there exist constants C and $c > 0$ with the following property. Let U be a random k-uniform hypergraph with vertex set $[n]$, and with each edge chosen independently with probability $p \ge Cn^{-1/k}$. Then with probability $1 - o(1)$ the following result holds. Let F be a subhypergraph of U that contains at most $cp^{k+1}n^{k+1}$ simplices. Then it is possible to remove at most apn^k edges from F and make it simplex free.*

Proof. Let $c = (1/8)(a/8)^{k+1}c_K$, where c_K is the constant given by Theorem 10.18, and apply that theorem to obtain a set E_F, which we shall take as our set E. Then E contains at most apn^k edges, so it remains to prove that when we remove the edges in E from F we obtain a hypergraph F' with no simplices.

Suppose we did have a simplex in F'. Let its vertex set be $\{x_1, \ldots, x_{k+1}\}$. For each $\{u, v\} \subset [k+1]$ of size 2, let J_{uv} be the set from the partition given by Theorem 10.18 that contains the $(k-1)$-set $\{x_i : i \notin \{u, v\}\}$. Then, as we commented before the statement of Theorem 10.17 (though then we were talking about H), for each u the set $F'(J_{uv} : v \in [k+1], v \ne u)$ is non-empty. Therefore, by Theorem 10.18, F', and hence F, contains at least $(1/4)(a/8)^{k+1}c_K p^{k+1}n^{k+1}$ simplices. By our choice of c, this is a contradiction.

This argument works for $Cn^{-1/k} \le p \le \delta$. However, since δ is a constant, we may, for $p > \delta$, simply apply the hypergraph removal lemma itself to remove all simplices. $\qquad\square$

62

10.4 The stability theorem

As a final application we will discuss the stability version of Turán's theorem, Theorem 1.11. The original stability theorem, due to Simonovits [57], is the following.

Theorem 10.20. *For every $\delta > 0$ and every graph H with $\chi(H) \geq 3$, there exists an $\epsilon > 0$ such that any H-free graph with at least $\left(1 - \frac{1}{\chi(H)-1} - \epsilon\right) \binom{n}{2}$ edges may be made $(\chi(H)-1)$-partite by removing at most δn^2 edges.*

Unfortunately, this is not quite enough for our purposes. We would like to be able to say that a graph that does not contain too many copies of H may be made $(\chi(H) - 1)$-partite by the deletion of few edges. To prove this, we appeal to the following generalization of the triangle removal lemma.

Theorem 10.21. *For every $\delta > 0$ and every graph H, there exists a constant $\epsilon > 0$ such that, if G is a graph containing at most ϵn^{v_H} copies of H, then it may be made H-free by removing at most δn^2 edges.*

Combining the two previous theorems gives us the robust statement we shall need.

Theorem 10.22. *For every $\delta > 0$ and every graph H with $\chi(H) \geq 3$, there exists a constant ϵ such that any graph with at most ϵn^{v_H} copies of H and at least $\left(1 - \frac{1}{\chi(H)-1} - \epsilon\right) \binom{n}{2}$ edges may be made $(\chi(H) - 1)$-partite by removing at most δn^2 edges.*

To prove Theorem 1.11, the statement of which we now repeat, we will follow the procedure described at the end of Section 3.

Theorem 10.23. *Given a strictly 2-balanced graph H with $\chi(H) \geq 3$ and a constant $\delta > 0$, there exist positive constants C and ϵ such that in the random graph $G_{n,p}$ chosen with probability $p \geq Cn^{-1/m_2(H)}$, where $m_2(H) = (e_H - 1)/(v_H - 2)$, the following holds with probability tending to 1 as n tends to infinity. Every H-free subgraph of $G_{n,p}$ with at least $\left(1 - \frac{1}{\chi(H)-1} - \epsilon\right) p\binom{n}{2}$ edges may be made $(\chi(H)-1)$-partite by removing at most $\delta p n^2$ edges.*

Proof. Fix $\delta > 0$. An application of Theorem 10.22 gives us $\epsilon > 0$ such that any graph with at most ϵn^{v_H} copies of H and at least $\left(1 - \frac{1}{\chi(H)-1} - 2\epsilon\right) \binom{n}{2}$ edges may be made $(\chi(H) - 1)$-partite by removing at most $\delta n^2/2$ edges.

Let $t = \chi(H) - 1$. Apply Corollary 9.7 with S the set of all labelled copies of H in K_n and $\mathcal{V}$ the set of all vertex subsets of $\{1, \ldots, n\}$. This yields constants C and λ such that, for $Cn^{-1/m_2(H)} \leq p \leq \lambda$, the following holds with probability tending to 1 as n tends to infinity. Let G be a random graph where each edge is chosen with probability p. Let μ be its characteristic measure. Then, if f is a function with $0 \leq f \leq \mu$, there exists a $\{0,1\}$-valued function j such that $\mathbb{E}_{s\in S} f(s_1) \ldots f(s_e) \geq \mathbb{E}_{s\in S} j(s_1) \cdots j(s_e) - \epsilon$ and, for all $V \in \mathcal{V}$, $|\mathbb{E}_{x\in V} f(x) - \mathbb{E}_{x\in V} j(x)| \leq \eta \frac{|X|}{|V|}$, where $\eta = \min(\epsilon, \delta/2t)$.

Let A be a H-free subgraph of G with $\left(1 - \frac{1}{\chi(H)-1} - \epsilon\right) p\binom{n}{2}$ edges and let $0 \leq f \leq \mu$ be p^{-1} times its characteristic function. Apply the transference principle to find the function j, which is the characteristic measure of a graph J. The number of copies of H in J is at most ϵn^{v_H}. Otherwise, we would have

$$\mathbb{E}_{s\in S} f(s_1) \ldots f(s_e) \geq \mathbb{E}_{s\in S} j(s_1) \cdots j(s_e) - \epsilon > 0,$$

implying that A was not H-free. Moreover, the number of edges in J is at least $\left(1 - \frac{1}{\chi(H)-1} - 2\epsilon\right)\binom{n}{2}$. Therefore, by the choice of ϵ, J may be made $(\chi(H)-1)$-partite by removing at most $\delta n^2/2$ edges.

Let $V_1, \ldots, V_t$ be the partite pieces. By transference, $|\mathbb{E}_{x \in V_i} f(x) - \mathbb{E}_{x \in V_i} j(x)| \leq \eta \frac{|X|}{|V_i|}$ for each $1 \leq i \leq t$. Therefore, if we remove all of the edges of A from each set in V_i, we have removed at most

$$\sum_{i=1}^{t} \sum_{x \in V_i} f(x) \leq \sum_{i=1}^{t} \sum_{x \in V_i} j(x) + t\eta|X| \leq \left(\frac{\delta}{2} + t\eta\right) n^2 \leq \delta n^2.$$

Moreover, the graph that remains is $(\chi(H)-1)$-partite, so we are done.

It only remains to consider the case when $p > \lambda$. However, as observed in [34], for p constant, the theorem follows from an application of the regularity lemma. This completes the proof. □

As a final note, we would like to mention that the method used in the proof of Theorem 10.23 should work quite generally. To take one more example, let K be the Fano plane. This is the hypergraph formed by taking the seven non-zero vectors of dimension three over the field with two elements and making xyz an edge if $x + y + z = 0$. The resulting graph has seven vertices and seven edges. It is known [4] that the extremal number of the Fano plane is approximately $\frac{3}{4}\binom{n}{3}$. Since the Fano plane is strictly 3-balanced, Theorem 10.9 implies that if U is a random 3-uniform hypergraph chosen with probability $p \geq Cn^{-2/3}$, then, with high probability, U is such that any subgraph of U with at least $\left(\frac{3}{4} + \epsilon\right)|U|$ edges contains the Fano plane.

Moreover, it was proved independently by Keevash and Sudakov [33] and Füredi and Simonovits [14] that the extremal example is formed by dividing the ground set into subsets A and B of nearly equal size and taking all triples that intersect both as edges. The stability version of this result says that, for all $\delta > 0$, there exists $\epsilon > 0$ such that any 3-uniform hypergraph on n vertices with at least $\left(\frac{3}{4} - \epsilon\right)\binom{n}{3}$ edges that does not contain the Fano plane may be partitioned into two parts A and B such that there are at most δn^3 edges contained entirely within A or B. The same proof as that of Theorem 10.23 then implies the following theorem.

Theorem 10.24. *Given a constant $\delta > 0$, there exist positive constants C and ϵ such that in the random graph $G_{n,p}^{(3)}$ chosen with probability $p \geq Cn^{-2/3}$, the following holds with probability tending to 1 as n tends to infinity. Every subgraph of $G_{n,p}^{(3)}$ with at least $\left(\frac{3}{4} - \epsilon\right)e(G)$ edges that does not contain the Fano plane may be made bipartite, in the sense that all edges intersect both parts of the partition, by removing at most δpn^3 edges.*

11 Concluding remarks

One question that the results of this paper leave open is to decide whether or not the thresholds we have proved are sharp. By saying that a threshold is sharp, we mean that the window over which the phase transition happens becomes arbitrarily small as the size of the ground set becomes large. For example, a graph property $\mathcal{P}$ has a sharp threshold at $\hat{p} = \hat{p}(n)$ if, for every $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}(G_{n,p} \text{ satisfies } \mathcal{P}) = \begin{cases} 0, & \text{if } p < (1-\epsilon)\hat{p}, \\ 1, & \text{if } p > (1+\epsilon)\hat{p}. \end{cases}$$

Connectedness is a simple example of a graph property for which a sharp threshold is known. The appearance of a triangle, on the other hand, is known not to be sharp. A result of Friedgut [10]

gives a criterion for judging whether a threshold is sharp or not. Roughly, this criterion says that if the property is globally determined, it is sharp, and if it is locally determined, it is not. This intuition allows one to conclude fairly quickly that connectedness should have a sharp threshold and the appearance of any particular small subgraph should not.

For the properties that we have discussed in this paper it is much less obvious whether the bounds are sharp or not. Many of the properties are not even monotone, which is crucial if one wishes to apply Friedgut's criterion. Nevertheless, the properties do not seem to be too pathological, so perhaps there is some small hope that the sharpness of their thresholds can be proved. There has even been some success in this direction already. Recall that the threshold at which $G_{n,p}$ becomes 2-colour Ramsey with respect to triangles is approximately $n^{-1/2}$. A difficult result of Friedgut, Rödl, Ruciński and Tetali [11] states that this threshold is sharp. That is, there exists $\hat{c} = \hat{c}(n)$ such that, for every $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}(G_{n,p} \text{ is } (K_3, 2)\text{-Ramsey}) = \begin{cases} 0, & \text{if } p < (1 - \epsilon)\hat{c}n^{-1/2}, \\ 1, & \text{if } p > (1 + \epsilon)\hat{c}n^{-1/2}. \end{cases}$$

Unfortunately, the function $\hat{c}(n)$ is not known to tend towards a constant. It could, at least in principle, wander up and down forever between the two endpoints. Nevertheless, we believe that extending this result to cover all (or any) of the theorems in this paper is important.

There are other improvements that it might well be possible to make. We proved our graph and hypergraph results for strictly balanced graphs and hypergraphs, while the results of Schacht [55] and Friedgut, Rödl and Schacht [12] apply to all graphs and hypergraphs. On the other hand, our methods allow us to prove structural results such as the stability theorem which do not seem to follow from their approach. It seems plausible that some synthesis of the two approaches could allow us to extend these latter results to general graphs and hypergraphs in a tidy fashion.[2]

In our approach, restricting to strictly balanced graphs and hypergraphs was very convenient, since it allowed us to cap our convolutions only at the very last stage (that is, when all the functions involved had sparse random support). In more general cases, capping would have to take place "all the way down". It seems likely that this can be done, but that a direct attempt to generalize our methods would be messy.

A more satisfactory approach would be to find a neater way of proving our probabilistic estimates. The process of capping is a bit ugly: a better approach might be to argue that with high probability we can say roughly how the modulus of an uncapped convolution is distributed, and use that in an inductive hypothesis. (It seems likely that the distribution is approximately Poisson.)

Thus, it seems that the problem of extending our methods to general graphs and hypergraphs and the problem of finding a neater proof of the probabilistic estimates go hand in hand.

References

[1] J. Balogh, R. Morris and W. Samotij, Independent sets in hypergraphs, to appear in *J. Amer. Math. Soc.*

---

[2]In subsequent work, Samotij [53] showed how to adapt Schacht's method so that it also applies to structural statements such as Theorem 1.11. However, it still remains an open problem to extend the methods of this paper to all graphs and hypergraphs.

[2] V. Bergelson and A. Leibman, Polynomial extensions of van der Waerden's and Szemerédi's theorems, *J. Amer. Math. Soc.* **9** (1996), 725–753.

[3] B. Bollobás, **Extremal Graph Theory**, Dover Publications, Inc., Mineola, NY, 2004. Reprint of the 1978 original.

[4] D. de Caen and Z. Füredi, The maximum size of 3-uniform hypergraphs not containing a Fano plane, *J. Combin. Theory Ser. B* **78** (2000), 274–276.

[5] D. Conlon, Combinatorial theorems relative to a random set, to appear in Proceedings of the 2014 ICM.

[6] D. Conlon, W. T. Gowers, W. Samotij and M. Schacht, On the KŁR conjecture in random graphs, to appear in *Israel J. Math.*

[7] P. Erdős, On the number of complete subgraphs contained in certain graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **7** (1962), 459–464.

[8] P. Erdős and M. Simonovits, Supersaturated graphs and hypergraphs, *Combinatorica* **3** (1983), 181–192.

[9] P. Frankl and V. Rödl, Large triangle-free subgraphs in graphs without K_4, *Graphs Combin.* **2** (1986), 135–144.

[10] E. Friedgut, Sharp thresholds of graph properties, and the k-sat problem, *J. Amer. Math. Soc.* **12** (1999), 1017–1054.

[11] E. Friedgut, V. Rödl, A. Ruciński and P. Tetali, A sharp threshold for random graphs with a monochromatic triangle in every edge coloring, *Mem. Amer. Math. Soc.* **179** (2006), no. 845, vi+66pp.

[12] E. Friedgut, V. Rödl and M. Schacht, Ramsey properties of discrete random structures, *Random Structures Algorithms* **37** (2010), 407–436.

[13] Z. Füredi, Random Ramsey graphs for the four-cycle, *Discrete Math.* **126** (1994), 407–410.

[14] Z. Füredi and M. Simonovits, Triple systems not containing a Fano configuration, *Combin. Probab. Comput.* **14** (2005), 467–484.

[15] H. Furstenberg and Y. Katznelson, An ergodic Szemerédi Theorem for commuting transformations, *J. Analyse Math.* **34** (1978), 275–291.

[16] S. Gerke, Y. Kohayakawa, V. Rödl and A. Steger, Small subsets inherit sparse ϵ-regularity, *J. Combin. Theory Ser. B* **97** (2007), 34–56.

[17] S. Gerke, H. J. Prömel, T. Schickinger, A. Steger and A. Taraz, K_4-free subgraphs of random graphs revisited, *Combinatorica* **27** (2007), 329–365.

[18] S. Gerke, A. Steger and T. Schickinger, K_5-free subgraphs of random graphs, *Random Structures Algorithms* **24** (2004), 194–232.

[19] W. T. Gowers, A new proof of Szemerédi's theorem, *Geom. Funct. Anal.* **11** (2001), 465–588.

[20] W. T. Gowers, Quasirandomness, counting and regularity for 3-uniform hypergraphs, *Combin. Probab. Comput.* **15** (2006), 143–184.

[21] W. T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, *Ann. of Math.* **166** (2007), 897–946.

[22] W. T. Gowers, Decompositions, approximate structure, transference, and the Hahn–Banach theorem, *Bull. London Math. Soc.* **42** (2010), 573–606.

[23] R. Graham, V. Rödl and A. Ruciński, On Schur properties of random subsets of integers, *J. Number Theory* **61** (1996), 388–408.

[24] B. Green and T. Tao, The primes contain arbitrarily long arithmetic progressions, *Ann. of Math.* **167** (2008), 481–547.

[25] M. Hamel and I. Łaba, Arithmetic structures in random sets, *Integers* **8** (2008), A04, 21pp.

[26] P. E. Haxell, Y. Kohayakawa and T. Łuczak, Turán's extremal problem in random graphs: forbidding even cycles, *J. Combin. Theory Ser. B* **64** (1995), 273–287.

[27] P. E. Haxell, Y. Kohayakawa and T. Łuczak, Turán's extremal problem in random graphs: forbidding odd cycles, *Combinatorica* **16** (1996), 107–122.

[28] S. Janson, Poisson approximation for large deviations, *Random Structures Algorithms* **1** (1990), 221–229.

[29] S. Janson, K. Oleszkiewicz and A. Ruciński, Upper tails for subgraph counts in random graphs, *Israel J. Math.* **142** (2004), 61–92.

[30] S. Janson and A. Ruciński, The infamous upper tail, *Random Structures Algorithms* **20** (2002), 317–342.

[31] S. Janson and A. Ruciński, The deletion method for upper tail estimates, *Combinatorica* **24** (2004), 615–640.

[32] S. Janson and A. Ruciński, Upper tails for counting objects in randomly induced subhypergraphs and rooted random graphs, *Ark. Mat.* **49** (2011), 79–96.

[33] P. Keevash and B. Sudakov, The Turán number of the Fano plane, *Combinatorica* **25** (2005), 561–574.

[34] Y. Kohayakawa, Szemerédi's regularity lemma for sparse graphs, in Foundations of computational mathematics (Rio de Janeiro, 1997), 216–230, Springer, Berlin, 1997.

[35] Y. Kohayakawa, T. Łuczak and V. Rödl, Arithmetic progressions of length three in subsets of a random set, *Acta Arith.* **75** (1996), 133–163.

[36] Y. Kohayakawa, T. Łuczak and V. Rödl, On K_4-free subgraphs of random graphs, *Combinatorica* **17** (1997), 173–213.

[37] Y. Kohayakawa, V. Rödl and M. Schacht, The Turán theorem for random graphs, *Combin. Probab. Comput.* **13** (2004), 61–91.

[38] T. Łuczak, Randomness and regularity, in International Congress of Mathematicians, Vol. III, 899–909, Eur. Math. Soc., Zürich, 2006.

[39] T. Łuczak, A. Ruciński and B.Voigt, Ramsey properties of random graphs, *J. Combin. Theory Ser. B* **56** (1992), 55–68.

[40] B. Nagle, V. Rödl and M. Schacht, The counting lemma for regular k-uniform hypergraphs, *Random Structures Algorithms* **28** (2006), 113–179.

[41] H. Nguyen, On two-point configurations in a random set, *Integers* **9** (2009), 41–45.

[42] R. Rado, Note on combinatorial analysis, *Proc. London Math. Soc.* **48** (1941), 122–160.

[43] F. P. Ramsey, On a problem of formal logic, *Proc. London Math. Soc.* **30** (1930), 264–286.

[44] O. Reingold, L. Trevisan, M. Tulsiani and S. Vadhan, Dense subsets of pseudorandom sets, in Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science, 76–85, IEEE Computer Society, Washington DC, 2008.

[45] V. Rödl and A. Ruciński, Lower bounds on probability thresholds for Ramsey properties, in Combinatorics, Paul Erdős is Eighty, Vol. 1, 317–346, Bolyai Soc. Math. Studies, János Bolyai Math. Soc., Budapest, 1993.

[46] V. Rödl and A. Ruciński, Random graphs with monochromatic triangles in every edge coloring, *Random Structures Algorithms* **5** (1994), 253–270.

[47] V. Rödl and A. Ruciński, Threshold functions for Ramsey properties, *J. Amer. Math. Soc.* **8** (1995), 917–942.

[48] V. Rödl and A. Ruciński, Rado partition theorem for random subsets of integers, *Proc. London Math. Soc.* **74** (1997), 481–502.

[49] V. Rödl and A. Ruciński, Ramsey properties of random hypergraphs, *J. Combin. Theory Ser. A* **81** (1998), 1–33.

[50] V. Rödl, A. Ruciński and M. Schacht, Ramsey properties of random k-partite, k-uniform hypergraphs, *SIAM J. Discrete Math.* **21** (2007), 442–460.

[51] V. Rödl and J. Skokan, Regularity lemma for uniform hypergraphs, *Random Structures Algorithms* **25** (2004), 1–42.

[52] K. F. Roth, On certain sets of integers, *J. London Math. Soc.* **28** (1953), 104–109.

[53] W. Samotij, Stability results for random discrete structures, *Random Structures Algorithms* **44** (2014), 269–289.

[54] D. Saxton and A. Thomason, Hypergraph containers, to appear in *Invent. Math.*

[55] M. Schacht, Extremal results for discrete random structures, *preprint*.

[56] I. Schur, Über die Kongruenz $x^m + y^m = z^m \, (\mathrm{mod} \ p)$, *Jber. Deutsch. Mat. Verein.* **25** (1916), 114–117.

[57] M. Simonovits, A method for solving extremal problems in graph theory, stability problems, in Theory of Graphs (Proc. Colloq., Tihany, 1966), 279–319, Academic Press, New York, 1968.

[58] T. Szabó and V. H. Vu, Turán's theorem in sparse random graphs, *Random Structures Algorithms* **23** (2003), 225–234.

[59] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 199–245.

[60] E. Szemerédi, Regular partitions of graphs, in Problèmes combinatoires et théorie des graphes (Orsay 1976), Colloq. Internat. CNRS, 260, 399–401, CNRS, Paris, 1978.

[61] T. Tao, A variant of the hypergraph removal lemma, *J. Combin. Theory Ser. A* **113** (2006), 1257–1280.

[62] Paul Turán, Egy gráfelméleti szélsőértékfeladatról, *Mat. Fiz. Lapok* **48** (1941), 436–452.

[63] B. L. van der Waerden, Beweis einer Baudetschen Vermutung, *Nieuw. Arch. Wisk.* **15** (1927), 212–216.

[64] P. Varnavides, On certain sets of positive density, *J. London Math. Soc.* **34** (1959), 358–360.

[65] V. H. Vu, A large deviation result on the number of small subgraphs of a random graph, *Combin. Probab. Comput.* **10** (2001), 79–94.